GenCore version 5.1.6

OM protein - protein search, using sw model

Run on: August 14, 2003, 13:03:37 ; Search time 94 Seconds
(without alignments)
494.143 Million cell updates/sec

Title: US-09-876-997-399
Perfect score: 968
Sequence: 1 MWLYRNPYVEAEYFPTKPMF.....RTCDYKHHWQDLLKCTNTAK 180

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 830525 seqs, 258052604 residues

Total number of hits satisfying chosen parameters: 830525

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SPTREMBL_23:*
1: sp_archea:*
2: sp_bacteria:*
3: sp_fungi:*
4: sp_human:*
5: sp_invertebrate:*
6: sp_mammal:*
7: sp_mhc:*
8: sp_organelle:*
9: sp_phage:*
10: sp_plant:*
11: sp_rodent:*
12: sp_virus:*
13: sp_vertebrate:*
14: sp_unclassified:*
15: sp_rvirus:*
16: sp_bacteriap:*
17: sp_archeap:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	926	95.7	223	4	Q8NEB5	Q8neb5 homo sapien
2	916	94.6	175	4	Q9BY45	Q9by45 homo sapien
3	440	45.5	363	5	Q8IGV3	Q8igv3 drosophila
4	440	45.5	412	5	Q9VND5	Q9vnd5 drosophila
5	382	39.5	314	10	Q9M882	Q9m882 arabidopsis
6	382	39.5	364	10	Q8LFD1	Q8lfd1 arabidopsis
7	379.5	39.2	317	10	Q8RZ97	Q8rz97 oryza sativ
8	379.5	39.2	328	10	Q94E07	Q94e07 oryza sativ
9	372	38.4	290	10	Q9XI60	Q9xi60 arabidopsis
10	364	37.6	374	10	Q9LLQ7	Q9llq7 vigna ungui
11	362	37.4	362	10	Q9AWT8	Q9awt8 oryza sativ
12	346	35.7	322	10	Q9FVL1	Q9fvl1 vigna ungui
13	337	34.8	302	10	Q8H0G3	Q8h0g3 arabidopsis
14	337	34.8	327	10	Q8LDP8	Q8ldp8 arabidopsis
15	337	34.8	327	10	Q9ZU49	Q9zu49 arabidopsis
16	336.5	34.8	308	10	Q8LAS9	Q8las9 arabidopsis
17	336	34.7	307	10	Q9LJQ8	Q9ljq8 arabidopsis
18	335	34.6	327	10	Q945N3	Q945n3 arabidopsis
19	247	25.5	289	3	Q05521	Q05521 saccharomyc
20	245	25.3	281	5	Q8IPQ4	Q8ipq4 drosophila
21	242	25.0	279	3	Q9UUA6	Q9uua6 schizosacch
22	205.5	21.2	328	10	Q8RZY0	Q8rzy0 oryza sativ
23	200.5	20.7	274	3	Q04396	Q04396 saccharomyc
24	196	20.2	335	5	Q8T2M2	Q8t2m2 dictyosteli
25	194.5	20.1	305	5	Q9VNT8	Q9vnt8 drosophila
26	188	19.4	340	5	Q8IA51	Q8ia51 caenorhabdi
27	188	19.4	346	5	Q9TXU1	Q9txu1 caenorhabdi
28	188	19.4	385	5	Q8IA52	Q8ia52 caenorhabdi
29	184.5	19.1	282	11	O08564	O08564 rattus norv
30	184.5	19.1	283	11	Q8K594	Q8k594 rattus norv
31	183.5	19.0	283	11	Q61469	Q61469 mus musculu
32	183.5	19.0	284	11	Q8BPB8	Q8bpb8 mus musculu
33	183	18.9	312	11	Q99JY8	Q99jy8 mus musculu
34	179.5	18.5	350	5	Q9NGZ2	Q9ngz2 drosophila
35	179.5	18.5	350	5	Q9V577	Q9v577 drosophila
36	179.5	18.5	350	5	Q8IGU7	Q8igu7 drosophila
37	179	18.5	312	11	P97544	P97544 rattus norv
38	178.5	18.4	288	4	O43688	O43688 homo sapien
39	174.5	18.0	286	11	O88957	O88957 cavia porce
40	173.5	17.9	285	11	O88956	O88956 cavia porce
41	173.5	17.9	311	4	Q96GW0	Q96gw0 homo sapien
42	173.5	17.9	311	4	O14495	O14495 homo sapien
43	172.5	17.8	284	4	O14494	O14494 homo sapien
44	172.5	17.8	285	4	O60457	O60457 homo sapien
45	172.5	17.8	289	4	O60463	O60463 homo sapien

ALIGNMENTS

RESULT 1
Q8NEB5
ID Q8NEB5 PRELIMINARY; PRT; 223 AA.
AC Q8NEB5;
DT 01-OCT-2002 (TrEMBLrel. 22, Created)
DT 01-OCT-2002 (TrEMBLrel. 22, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE Similar to HTPAP protein.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Testis;
RA Strausberg R.;
RL Submitted (JUN-2002) to the EMBL/GenBank/DDBJ databases.
DR EMBL; BC033025; AAH33025.1; -.
DR InterPro; IPR000326; PA_PTPase.
DR Pfam; PF01569; PAP2; 1.
DR SMART; SM00014; acidPPc; 1.
SQ SEQUENCE 223 AA; 25159 MW; 72B01C9A0DBBA1BA CRC64;

Query Match 95.7%; Score 926; DB 4; Length 223;
Best Local Similarity 99.4%; Pred. No. 2.2e-94;
Matches 172; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

```
Qy          1 MWLYRNPYVEAEYFPTKPMFVIAFLSPLSLIFLAKFLKKADTRDSRQACLAASLALALNG 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MWLYRNPYVEAEYFPTKPMFVIAFLSPLSLIFLAKFLKKADTRDSRQACLAASLALALNG 60

Qy         61 VFTNTIKLIVGRPRPDFFYRCFPDGLAHSDLMCTGDKDVVNEGRKSFPSGHSSFAFAGLA 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 VFTNTIKLIVGRPRPDFFYRCFPDGLAHSDLMCTGDKDVVNEGRKSFPSGHSSFAFAGLA 120

Qy        121 FASFYLAGKLHCFTPQGRGKSWRFCAFLSPLLFAAVIALSRTCDYKHHWQDLL 173
              |||||||||||||||||||||||||||||||||||||||||||||||||||:|
Db        121 FASFYLAGKLHCFTPQGRGKSWRFCAFLSPLLFAAVIALSRTCDYKHHWQDVL 173
```

```
RESULT 2
Q9BY45
ID   Q9BY45      PRELIMINARY;      PRT;   175 AA.
AC   Q9BY45;
DT   01-JUN-2001 (TrEMBLrel. 17, Created)
DT   01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DT   01-OCT-2001 (TrEMBLrel. 18, Last annotation update)
DE   HTPAP.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Liver;
RA   Li Y., Wu T., Xu S., Ren S., Chen Z., Han Z.;
RT   "A novel gene expressed in human liver non-tumor tissue.";
RL   Submitted (DEC-1999) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AF212238; AAK14924.1; -.
DR   InterPro; IPR000326; PA_PTPase.
DR   Pfam; PF01569; PAP2; 1.
DR   SMART; SM00014; acidPPc; 1.
SQ   SEQUENCE   175 AA;  19766 MW;  0133956D40539F83 CRC64;

  Query Match             94.6%;  Score 916;  DB 4;  Length 175;
  Best Local Similarity   98.3%;  Pred. No. 2.2e-93;
  Matches  171;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;

Qy          1 MWLYRNPYVEAEYFPTKPMFVIAFLSPLSLIFLAKFLKKADTRDSRQACLAASLALALNG 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MWLYRNPYVEAEYFPTKPMFVIAFLSPLSLIFLAKFLKKADTRDSRQACLAASLALALNG 60

Qy         61 VFTNTIKLIVGRPRPDFFYRCFPDGLAHSDLMCTGDKDVVNEGRKSFPSGHSSFAFAGLA 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 VFTNTIKLIVGRPRPDFFYRCFPDGLAHSDLMCTGDKDVVNEGRKSFPSGHSSFAFAGLA 120

Qy        121 FASFYLAGKLHCFTPQGRGKSWRFCAFLSPLLFAAVIALSRTCDYKHHWQDLLK 174
              ||||||||||||||||||||||||||||||||||||||||||||||||||    |
Db        121 FASFYLAGKLHCFTPQGRGKSWRFCAFLSPLLFAAVIALSRTCDYKHHWQGPFK 174


RESULT 3
Q8IGV3
ID   Q8IGV3      PRELIMINARY;      PRT;   363 AA.
AC   Q8IGV3;
DT   01-MAR-2003 (TrEMBLrel. 23, Created)
DT   01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DT   01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE   RE23632p.
GN   CG12746.
OS   Drosophila melanogaster (Fruit fly).
OC   Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC   Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC   Ephydroidea; Drosophilidae; Drosophila.
OX   NCBI_TaxID=7227;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=y;
RA   Stapleton M., Brokstein P., Hong L., Agbayani A., Carlson J.,
RA   Champe M., Chavez C., Dorsett V., Dresnek D., Farfan D., Frise E.,
RA   George R., Gonzalez M., Guarin H., Kronmiller B., Li P., Liao G.,
RA   Miranda A., Mungall C.J., Nunoo J., Pacleb J., Paragas V., Park S.,
RA   Patel S., Phouanenavong S., Wan K., Yu C., Lewis S.E., Rubin G.M.,
RA   Celniker S.;
RL   Submitted (NOV-2002) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; BT001579; AAN71334.1; -.
SQ   SEQUENCE   363 AA;  41230 MW;  D46A8483FD4F8E72 CRC64;

  Query Match             45.5%;  Score 440;  DB 5;  Length 363;
  Best Local Similarity   44.5%;  Pred. No. 2.8e-40;
  Matches   85;  Conservative   29;  Mismatches   49;  Indels   28;  Gaps    4;

Qy          1 MWLYRNPYVEAEYFPTKP-------MFVIAFLSPLSLIFLAKFLKKADTRDSRQACLAAS 53
              :|||:|         | :|       :     ::|  |: :| :    | || | |  | :
Db        103 LWLYKN--------PRRPDIVRGGELLFWVIVAPF-LVTIAFYWYTRDRRDFRAASWAWT 153

Qy         54 LALALNGVFTNTIKLIVGRPRPDFFYRCFPDGL-----------AHSDLMCTGDKDVVN 101
              ||| :||: |: :|: |||||||:||||||||:           :  |  |||   :|
Db        154 LALCMNGIPTSVLKITVGRPRPDYFYRCFPDGVMVLNTTSNGVDTSILDFNCTGLPGDIN 213

Qy        102 EGRKSFPSGHSSFAFAGLAFASFYLAGKLHCFTPQGRGKSWRFCAFLSPLLFAAVIALSR 161
              |||||||||||||||   | ::|:  ||| |  :||| :|| |  : ||  | ::|:||
Db        214 EGRKSFPSGHSSFAFASFGFIAYYIGAKLHAFDSRGRGHTWRLCIAVIPLFIALLVAVSR 273

Qy        162 TCDYKHHWQDL 172
              |||| |||||:
Db        274 TCDYHHHWQDV 284


RESULT 4
Q9VND5
ID   Q9VND5      PRELIMINARY;      PRT;   412 AA.
AC   Q9VND5; Q9VND6;
DT   01-MAY-2000 (TrEMBLrel. 13, Created)
DT   01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT   01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE   CG12746 protein.
GN   CG12746.
OS   Drosophila melanogaster (Fruit fly).
OC   Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC   Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC   Ephydroidea; Drosophilidae; Drosophila.
OX   NCBI_TaxID=7227;
RN   [1]
RP   SEQUENCE FROM N.A. (LONG AND SHORT ISOFORMS).
RC   STRAIN=BERKELEY;
RX   MEDLINE=20196006; PubMed=10731132;
RA   Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA   Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
RA   George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA   Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA   Brandon R.C., Rogers Y.-H.C., Blazej R.G., Champe M., Pfeiffer B.D.,
RA   Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
RA   Abril J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
RA   Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA   Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
RA   Borkova D., Botchan M.R., Bouck J., Brokstein P., Brottier P.,
RA   Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA   Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA   de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA   Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA   Durbin K.J., Evangelista C.C., Ferraz C., Ferriera S., Fleischmann W.,
RA   Fosler C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,
RA   Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA   Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
RA   Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
RA   Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA   Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA   Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA   Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
RA   Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
RA   Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA   Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleb J.M.,
RA   Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA   Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
RA   Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
RA   Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA   Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA   Wang Z.-Y., Wassarman D.A., Weinstock G.M., Weissenbach J.,
RA   Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
RA   Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA   Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA   Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RT   "The genome sequence of Drosophila melanogaster.";
RL   Science 287:2185-2195(2000).
```

```
CC   -!- ALTERNATIVE PRODUCTS: 2 ISOFORMS; A LONG FORM (SHOWN HERE) AND A
CC       SHORT FORM; ARE PRODUCED BY ALTERNATIVE SPLICING.
DR   EMBL; AE003602; AAF52007.1; -.
DR   EMBL; AE003602; AAF52008.1; -.
DR   FlyBase; FBgn0037341; CG12746.
DR   InterPro; IPR000326; PA_PTPase.
DR   Pfam; PF01569; PAP2; 1.
DR   SMART; SM00014; acidPPc; 1.
KW   Alternative splicing; Hypothetical protein.
FT   VARSPLIC      1     49       MISSING (IN SHORT ISOFORM).
SQ   SEQUENCE   412 AA;  46736 MW;  0966C4340A4444C1 CRC64;

  Query Match             45.5%;  Score 440;  DB 5;  Length 412;
  Best Local Similarity   44.5%;  Pred. No. 3.2e-40;
  Matches   85;  Conservative   29;  Mismatches   49;  Indels   28;  Gaps    4;

Qy          1 MWLYRNPYVEAEYFPTKP-------MFVIAFLSPLSLIFLAKFLKKADTRDSRQACLAAS 53
              :|||:|        | :|       :     ::|  |: :| :    | || | |  | :
Db        152 LWLYKN--------PRRPDIVRGGELLFWVIVAPF-LVTIAFYWYTRDRRDFRAASWAWT 202

Qy         54 LALALNGVFTNTIKLIVGRPRPDFFYRCFPDGL------------AHSDLMCTGDKDVVN 101
              ||| :||: |: :|: ||||||:||||||||:            :  |  |||    :|
Db        203 LALCMNGIPTSVLKITVGRPRPDYFYRCFPDGVMVLNTTSNGVDTSILDFNCTGLPGDIN 262

Qy        102 EGRKSFPSGHSSFAFAGLAFASFYLAGKLHCFTPQGRGKSWRFCAFLSPLLFAAVIALSR 161
              |||||||||||||||   | ::|:  ||| |  :||| :|| |  : ||  | ::|:||
Db        263 EGRKSFPSGHSSFAFASFGFIAYYIGAKLHAFDSRGRGHTWRLCIAVIPLFIALLVAVSR 322

Qy        162 TCDYKHHWQDL 172
              |||| |||||:
Db        323 TCDYHHHWQDV 333


RESULT 5
Q9M882
ID   Q9M882      PRELIMINARY;      PRT;   314 AA.
AC   Q9M882;
DT   01-OCT-2000 (TrEMBLrel. 15, Created)
DT   01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT   01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE   Putative phosphatidate phosphohydrolase.
GN   F16B3.23.
OS   Arabidopsis thaliana (Mouse-ear cress).
OC   Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC   Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC   eurosids II; Brassicales; Brassicaceae; Arabidopsis.
OX   NCBI_TaxID=3702;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=cv. Columbia;
RA   Lin X., Kaul S., Town C.D., Benito M., Creasy T.H., Haas B., Wu D.,
RA   Ronning C.M., Koo H., Fujii C.Y., Utterback T.R., Barnstead M.E.,
RA   Bowman C.L., White O., Nierman W.C., Fraser C.M.;
RT   "Arabidopsis thaliana chromosome III BAC F16B3 genomic sequence.";
RL   Submitted (JAN-2001) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AC021640; AAF32467.1; -.
DR   InterPro; IPR000326; PA_PTPase.
DR   Pfam; PF01569; PAP2; 1.
DR   SMART; SM00014; acidPPc; 1.
KW   Hydrolase.
SQ   SEQUENCE   314 AA;  35184 MW;  5F1E88546058C497 CRC64;

  Query Match             39.5%;  Score 382;  DB 10;  Length 314;
  Best Local Similarity   47.0%;  Pred. No. 6.3e-34;
  Matches   77;  Conservative   28;  Mismatches   49;  Indels   10;  Gaps    4;

Qy         16 TKPMF---VIAFLSPLSLIFLAKFLKKADTRDSRQACLAASLALALNGVFTNTIKLIVGR 72
              | |::   | | | || :||:  : :: |  |   | |    :: :  | |: ||  |||
Db         65 TVPIWSVPVYAMLLPL-VIFIFIYFRRRDVYDLHHAVLGLLYSVLVTAVLTDAIKNAVGR 123

Qy         73 PRPDFFYRCFPDGLA----HSDLMCTGDKDVVNEGRKSFPSGHSSFAFAGLAFASFYLAG 128
              ||||||:|||||| |      |::| ||| |: || |||||||:|::|::|| | | ||:|
Db        124 PRPDFFWRCFPDGKALYDSLGDVICHGDKSVIREGHKSFPSGHTSWSFSGLGFLSLYLSG 183

Qy        129 KLHCFTPQGRGKSWRFCAFLSPLLFAAVIALSRTCDYKHHWQDL 172
              |:  |   |:|   : |  : ||||||:: :||  || |||||:
Db        184 KIQAF--DGKGHVAKLCIVILPLLFAALVGISRVDDYWHHWQDV 225


RESULT 6
Q8LFD1
ID   Q8LFD1      PRELIMINARY;      PRT;   364 AA.
AC   Q8LFD1;
DT   01-OCT-2002 (TrEMBLrel. 22, Created)
DT   01-OCT-2002 (TrEMBLrel. 22, Last sequence update)
DT   01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE   Putative phosphatidate phosphohydrolase.
OS   Arabidopsis thaliana (Mouse-ear cress).
OC   Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC   Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC   eurosids II; Brassicales; Brassicaceae; Arabidopsis.
OX   NCBI_TaxID=3702;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Haas B.J., Volfovsky N., Town C.D., Troukhan M., Alexandrov N.,
RA   Feldmann K.A., Flavell R.B., White O., Salzberg S.L.;
RT   "Full-length messenger RNA sequences greatly improve genome
RT   annotation.";
RL   Genome Biol. 0:0-0(2002).
RN   [2]
RP   SEQUENCE FROM N.A.
RA   Brover V., Troukhan M., Alexandrov N., Lu Y.-P., Flavell R.,
RA   Feldmann K.;
RT   "Full-Length cDNA from Arabidopsis thaliana.";
RL   Submitted (MAR-2002) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AY084915; AAM61477.1; -.
DR   InterPro; IPR000326; PA_PTPase.
DR   Pfam; PF01569; PAP2; 1.
DR   SMART; SM00014; acidPPc; 1.
KW   Hydrolase.
SQ   SEQUENCE   364 AA;  40775 MW;  905F149C235E4A18 CRC64;

  Query Match             39.5%;  Score 382;  DB 10;  Length 364;
  Best Local Similarity   47.0%;  Pred. No. 7.5e-34;
  Matches   77;  Conservative   28;  Mismatches   49;  Indels   10;  Gaps    4;

Qy         16 TKPMF---VIAFLSPLSLIFLAKFLKKADTRDSRQACLAASLALALNGVFTNTIKLIVGR 72
              | |::   | | | || :||:  : :: |  |   | |    :: :  | |: ||  |||
Db        115 TVPIWSVPVYAMLLPL-VIFIFIYFRRRDVYDLHHAVLGLLYSVLVTAVLTDAIKNAVGR 173

Qy         73 PRPDFFYRCFPDGLA----HSDLMCTGDKDVVNEGRKSFPSGHSSFAFAGLAFASFYLAG 128
              ||||||:|||||| |      |::| ||| |: || |||||||:|::|::|| | | ||:|
Db        174 PRPDFFWRCFPDGKALYDSLGDVICHGDKSVIREGHKSFPSGHTSWSFSGLGFLSLYLSG 233

Qy        129 KLHCFTPQGRGKSWRFCAFLSPLLFAAVIALSRTCDYKHHWQDL 172
              |:  |   |:|   : |  : ||||||:: :||  || |||||:
Db        234 KIQAF--DGKGHVAKLCIVILPLLFAALVGISRVDDYWHHWQDV 275


RESULT 7
Q8RZ97
ID   Q8RZ97      PRELIMINARY;      PRT;   317 AA.
AC   Q8RZ97;
DT   01-JUN-2002 (TrEMBLrel. 21, Created)
DT   01-JUN-2002 (TrEMBLrel. 21, Last sequence update)
DT   01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
DE   Putative phosphatidic acid phosphatase beta.
GN   OSJNBB0063G05.1.
OS   Oryza sativa (japonica cultivar-group).
OC   Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC   Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
OC   Ehrhartoideae; Oryzeae; Oryza.
OX   NCBI_TaxID=39947;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=cv. Nipponbare;
```

```
RA   Sasaki T., Matsumoto T., Yamamoto K.;
RT   "Oryza sativa nipponbare(GA3) genomic DNA, chromosome 1, BAC
RT   clone:OSJNBb0063G05.";
RL   Submitted (JUN-2001) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AP003760; BAB90492.1; -.
DR   Gramene; Q8RZ97; -.
DR   InterPro; IPR000326; PA_PTPase.
DR   Pfam; PF01569; PAP2; 1.
DR   SMART; SM00014; acidPPc; 1.
SQ   SEQUENCE   317 AA;  34832 MW;  9008B4DF7FC5D849 CRC64;

  Query Match             39.2%;  Score 379.5;  DB 10;  Length 317;
  Best Local Similarity   45.0%;  Pred. No. 1.2e-33;
  Matches   76;  Conservative   33;  Mismatches   47;  Indels   13;  Gaps    5;

Qy         16 TKPMF---VIAFLSPLSLIFLAKFLKKADTRDSRQACLAASLALALNGVFTNTIKLIVGR 72
              | |::   ::| : |: ::|   : :: :  |   | |    :: : || |: ||  |||
Db         71 TVPIWAVPIVAVIGPM-IVFTVVYFRRRNVYDLHHAVLGILFSVLITGVLTDAIKDAVGR 129

Qy         73 PRPDFFYRCFPDGLAHSD-----LMCTGDKDVVNEGRKSFPSGHSSFAFAGLAFASFYLA 127
              |||:||:||||||:|  |      ::| ||  |: || |||||||:|::|||| | |:|||
Db        130 PRPNFFWRCFPDGIAVFDNVTTGVICHGDASVIKEGHKSFPSGHTSWSFAGLGFLSWYLA 189

Qy        128 GKLHCFTPQGRGKSWRFCAFLSPLLFAAVIALSRTCDYKHHWQDLLKCT 176
              ||:  |    ||   : |  |:||| ||::|:||  || |||||:  ||
Db        190 GKITVF--DRRGHVAKLCVVLAPLLVAAMVAISRVDDYWHHWQDV--CT 234


RESULT 8
Q94E07
ID   Q94E07      PRELIMINARY;      PRT;   328 AA.
AC   Q94E07;
DT   01-DEC-2001 (TrEMBLrel. 19, Created)
DT   01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT   01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE   Putative phosphatidic acid phosphatase beta.
GN   P0003E08.23.
OS   Oryza sativa (Rice).
OC   Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC   Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
OC   Ehrhartoideae; Oryzeae; Oryza.
OX   NCBI_TaxID=4530;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=cv. Nipponbare;
RA   Sasaki T., Matsumoto T., Yamamoto K.;
RT   "Oryza sativa nipponbare(GA3) genomic DNA, chromosome 1, PAC
RT   clone:P0003E08.";
RL   Submitted (FEB-2001) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AP003222; BAB63556.1; -.
DR   Gramene; Q94E07; -.
DR   InterPro; IPR000326; PA_PTPase.
DR   Pfam; PF01569; PAP2; 1.
DR   SMART; SM00014; acidPPc; 1.
SQ   SEQUENCE   328 AA;  36004 MW;  5CDD3A44ABDB9C55 CRC64;

  Query Match             39.2%;  Score 379.5;  DB 10;  Length 328;
  Best Local Similarity   45.0%;  Pred. No. 1.3e-33;
  Matches   76;  Conservative   33;  Mismatches   47;  Indels   13;  Gaps    5;

Qy         16 TKPMF---VIAFLSPLSLIFLAKFLKKADTRDSRQACLAASLALALNGVFTNTIKLIVGR 72
              | |::   ::| : |: ::|   : :: :  |   | |    :: : || |: ||  |||
Db         71 TVPIWAVPIVAVIGPM-IVFTVVYFRRRNVYDLHHAVLGILFSVLITGVLTDAIKDAVGR 129

Qy         73 PRPDFFYRCFPDGLAHSD-----LMCTGDKDVVNEGRKSFPSGHSSFAFAGLAFASFYLA 127
              |||:||:||||||:|  |      ::| ||  |: || |||||||:|::|||| | |:|||
Db        130 PRPNFFWRCFPDGIAVFDNVTTGVICHGDASVIKEGHKSFPSGHTSWSFAGLGFLSWYLA 189

Qy        128 GKLHCFTPQGRGKSWRFCAFLSPLLFAAVIALSRTCDYKHHWQDLLKCT 176
              ||:  |    ||   : |  |:||| ||::|:||  || |||||:  ||
Db        190 GKITVF--DRRGHVAKLCVVLAPLLVAAMVAISRVDDYWHHWQDV--CT 234


RESULT 9
Q9XI60
ID   Q9XI60      PRELIMINARY;      PRT;   290 AA.
AC   Q9XI60;
DT   01-NOV-1999 (TrEMBLrel. 12, Created)
DT   01-NOV-1999 (TrEMBLrel. 12, Last sequence update)
DT   01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE   F9L1.2 protein (Putative phosphatidic acid phosphatase) (Prenyl
DE   diphosphate phosphatase).
GN   F9L1.2 OR ATPAP2.
OS   Arabidopsis thaliana (Mouse-ear cress).
OC   Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC   Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC   eurosids II; Brassicales; Brassicaceae; Arabidopsis.
OX   NCBI_TaxID=3702;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=cv. columbia;
RA   Vysotskaia V.S., Schwartz J.R., Yu G., Toriumi M., Lenz C., Liu S.,
RA   Lee J., Liu A., Li J., Kremenetskaia I., Luros J., Gonzalez A.,
RA   Altafi H., Araujo R., Brooks S., Buehler E., Chao Q., Conn L.,
RA   Conway A.B., Dunn P., Hansen N., Huizar L., Khan S., Kim C., Palm C.,
RA   Rowley D., Shinn P., Walker M., Davis R.W., Ecker J.R.,
RA   Federspiel N.A., Theologis A.;
RT   "Arabidopsis thaliana chromosome 1 BAC F9L1 sequence.";
RL   Submitted (JUN-1999) to the EMBL/GenBank/DDBJ databases.
RN   [2]
RP   SEQUENCE FROM N.A.
RA   Katagiri T., Shinozaki K.;
RT   "A cDNA sequence encoding a phosphatidic acid phosphatase in
RT   Arabidopsis thaliana.";
RL   Submitted (MAY-2001) to the EMBL/GenBank/DDBJ databases.
RN   [3]
RP   SEQUENCE FROM N.A.
RA   Haas B.J., Volfovsky N., Town C.D., Troukhan M., Alexandrov N.,
RA   Feldmann K.A., Flavell R.B., White O., Salzberg S.L.;
RT   "Full-length messenger RNA sequences greatly improve genome
RT   annotation.";
RL   Genome Biol. 0:0-0(2002).
RN   [4]
RP   SEQUENCE FROM N.A.
RA   Brover V., Troukhan M., Alexandrov N., Lu Y.-P., Flavell R.,
RA   Feldmann K.;
RT   "Full-Length cDNA from Arabidopsis thaliana.";
RL   Submitted (MAR-2002) to the EMBL/GenBank/DDBJ databases.
RN   [5]
RP   SEQUENCE FROM N.A.
RC   STRAIN=cv. Columbia;
RA   Tokuhiro K., Muramoto N., Yamada Y., Asami O., Hirai M., Obata S.,
RA   Ohto C., Muramatsu M.;
RT   "Prenyl alcohol production by overexpression of prenyl diphosphate
RT   phosphatase in Yeast Saccharomyces cerevisiae.";
RL   Submitted (JAN-2001) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AC007591; AAD39637.1; -.
DR   EMBL; AB061407; BAB47574.1; -.
DR   EMBL; AY087673; AAM65210.1; -.
DR   EMBL; AB053950; BAC41335.1; -.
DR   InterPro; IPR000326; PA_PTPase.
DR   Pfam; PF01569; PAP2; 1.
SQ   SEQUENCE   290 AA;  32702 MW;  BF14A9A0C23B4429 CRC64;

  Query Match             38.4%;  Score 372;  DB 10;  Length 290;
  Best Local Similarity   47.8%;  Pred. No. 7.4e-33;
  Matches   75;  Conservative   28;  Mismatches   46;  Indels    8;  Gaps    4;

Qy         21 VIAFLSPLSLIFLAKFLKKADTRDSRQACLAASLALALNGVFTNTIKLIVGRPRPDFFYR 80
              :|| : | ::| :  |::  |  |    | |    :: : || |: ||  |||||||||:|
Db         73 LIAVVLPFAVICVYYFIRN-DVYDLHHAILGLLFSVLITGVITDAIKDAVGRPRPDFFWR 131

Qy         81 CFPDGLA--HS---DLMCTGDKDVVNEGRKSFPSGHSSFAFAGLAFASFYLAGKLHCFTP 135
              |||||:   |:   :::||| |||| || |||||||:|::|||| | | ||:||:  |
Db        132 CFPDGIGIFHNVTKNVLCTGAKDVVKEGHKSFPSGHTSWSFAGLGFLSLYLSGKIRVF-- 189
```

```
Qy        136 QGRGKSWRFCAFLSPLLFAAVIALSRTCDYKHHWQDL 172
              ||   : |  : ||| ||:: :||  || |||||:
Db        190 DQRGHVAKLCIVILPLLVAALVGVSRVDDYWHHWQDV 226


RESULT 10
Q9LLQ7
ID   Q9LLQ7      PRELIMINARY;      PRT;   374 AA.
AC   Q9LLQ7;
DT   01-OCT-2000 (TrEMBLrel. 15, Created)
DT   01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT   01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE   Phosphatidic acid phosphatase alpha (EC 3.1.3.4).
OS   Vigna unguiculata (Cowpea).
OC   Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC   Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC   eurosids I; Fabales; Fabaceae; Papilionoideae; Phaseoleae; Vigna.
OX   NCBI_TaxID=3917;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=cv. EPACE-1; TISSUE=Leaf;
RA   Franca M.G.C., Matos A.R., d'Arcy-Lameta A., Zuily-Fodil Y.,
RA   Pham-Thi A.T.;
RL   Submitted (JUL-1999) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AF165891; AAF89579.1; -.
DR   InterPro; IPR000326; PA_PTPase.
DR   Pfam; PF01569; PAP2; 1.
DR   SMART; SM00014; acidPPc; 1.
KW   Hydrolase.
SQ   SEQUENCE   374 AA;  42336 MW;  B50DAD5707D0A0BC CRC64;

  Query Match             37.6%;  Score 364;  DB 10;  Length 374;
  Best Local Similarity   45.7%;  Pred. No. 7.6e-32;
  Matches   75;  Conservative   27;  Mismatches   52;  Indels   10;  Gaps    4;

Qy         16 TKPMF---VIAFLSPLSLIFLAKFLKKADTRDSRQACLAASLALALNGVFTNTIKLIVGR 72
              | |::   : | | |: :|||  :::: |  |  | |    :| :  | |  ||  |||
Db        120 TVPVWSIPIYAVLLPI-VIFLVVYIRRRDVYDLHHAVLGLLFSLLITAVITEAIKNGVGR 178

Qy         73 PRPDFFYRCFPDGL----AHSDLMCTGDKDVVNEGRKSFPSGHSSFAFAGLAFASFYLAG 128
              ||||||:||||||        |::| |:| || || |||||||:|::|:|| | | ||:|
Db        179 PRPDFFWRCFPDGKDVYDKLGDVICHGEKGVVKEGYKSFPSGHTSWSFSGLGFLSLYLSG 238

Qy        129 KLHCFTPQGRGKSWRFCAFLSPLLFAAVIALSRTCDYKHHWQDL 172
              |:  |    ||   : |    |||||::: :||  || |||||:
Db        239 KIKAF--DRRGHVAKLCIVFLPLLFASLVGISRVDDYWHHWQDV 280


RESULT 11
Q9AWT8
ID   Q9AWT8      PRELIMINARY;      PRT;   362 AA.
AC   Q9AWT8;
DT   01-JUN-2001 (TrEMBLrel. 17, Created)
DT   01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DT   01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
DE   Putative phosphatidic acid phosphatase alpha.
GN   P0480E02.6.
OS   Oryza sativa (Rice).
OC   Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC   Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
OC   Ehrhartoideae; Oryzeae; Oryza.
OX   NCBI_TaxID=4530;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=cv. Nipponbare;
RA   Sasaki T., Matsumoto T., Yamamoto K.;
RT   "Oryza sativa nipponbare(GA3) genomic DNA, chromosome 1, PAC
RT   clone:P0480E02.";
RL   Submitted (NOV-2000) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AP002913; BAB21200.1; -.
DR   Gramene; Q9AWT8; -.
DR   InterPro; IPR000326; PA_PTPase.
DR   Pfam; PF01569; PAP2; 1.
DR   SMART; SM00014; acidPPc; 1.
SQ   SEQUENCE   362 AA;  40682 MW;  8C4F1141F8BF5176 CRC64;

  Query Match             37.4%;  Score 362;  DB 10;  Length 362;
  Best Local Similarity   40.1%;  Pred. No. 1.2e-31;
  Matches   71;  Conservative   36;  Mismatches   62;  Indels    8;  Gaps    3;

Qy          1 MWLYRNPYVEAEYFPTKPMFVIAFLSPLSLIFLAKFLKKADTRDSRQACLAASLALALNG 60
              |  | :  |:     |    : :|: : |: :|||: :: : |  |   | |    :: :
Db         56 MMTYVSYPVKQSTVPAWGVPIISIVCPV-IIFLSVYIARRDVYDLHHATLGVLYSVLITA 114

Qy         61 VFTNTIKLIVGRPRPDFFYRCFPDG-----LAHSDLMCTGDKDVVNEGRKSFPSGHSSFA 115
              | |  :|  |||||||||:||||||         |::| |:|  : :||||||||:|::
Db        115 VVTTVVKNAVGRPRPDFFWRCFPDGKQLYDQVTGDVICHGEKSFLKDGRKSFPSGHTSWS 174

Qy        116 FAGLAFASFYLAGKLHCFTPQGRGKSWRFCAFLSPLLFAAVIALSRTCDYKHHWQDL 172
              |||| | | ||:||:  |  ||     : |  : ||| |::: :||  ||:|||:|:
Db        175 FAGLGFLSLYLSGKIKVFDRQGHVA--KLCIMILPLLIASLVGISRIDDYRHHWEDV 229


RESULT 12
Q9FVL1
ID   Q9FVL1      PRELIMINARY;      PRT;   322 AA.
AC   Q9FVL1;
DT   01-MAR-2001 (TrEMBLrel. 16, Created)
DT   01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT   01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE   Phosphatidic acid phosphatase beta (EC 3.1.3.4).
OS   Vigna unguiculata (Cowpea).
OC   Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC   Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC   eurosids I; Fabales; Fabaceae; Papilionoideae; Phaseoleae; Vigna.
OX   NCBI_TaxID=3917;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=cv. EPACE-1; TISSUE=Leaf;
RA   Franca M.G.C., Matos A.R., d'Arcy-Lameta A., Zuily-Fodil Y.,
RA   Pham-Thi A.T.;
RL   Submitted (OCT-2000) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AF171230; AAF89745.3; -.
DR   InterPro; IPR000326; PA_PTPase.
DR   Pfam; PF01569; PAP2; 1.
DR   SMART; SM00014; acidPPc; 1.
KW   Hydrolase.
SQ   SEQUENCE   322 AA;  36224 MW;  67B7B572DCAA1F8E CRC64;

  Query Match             35.7%;  Score 346;  DB 10;  Length 322;
  Best Local Similarity   45.9%;  Pred. No. 6.4e-30;
  Matches   72;  Conservative   25;  Mismatches   52;  Indels    8;  Gaps    3;

Qy         21 VIAFLSPLSLIFLAKFLKKADTRDSRQACLAASLALALNGVFTNTIKLIVGRPRPDFFYR 80
              ::| | ||::  :  |::| |  |   | |    :: :  | |: ||  ||||||||:|
Db         73 IVAILLPLAVFLVYYFIRK-DVYDFHHAILGLLFSVLITAVITDAIKDGVGRPRPDFFWR 131

Qy         81 CFPDGL-----AHSDLMCTGDKDVVNEGRKSFPSGHSSFAFAGLAFASFYLAGKLHCFTP 135
              |||||        ||: ||||| |: || |  ||||:|::|||| : |: |:||:  |
Db        132 CFPDGKGVFDPVTSDVRCTGDKGVIKEGTKVSPSGHTSWSFAGLVYLSWKLSGKIRVF-- 189

Qy        136 QGRGKSWRFCAFLSPLLFAAVIALSRTCDYKHHWQDL 172
                ||   : |    |:| ||:|| ||  || |||||:
Db        190 DRRGHVAKLCLVFLPILVAAMIAGSRVDDYWHHWQDV 226


RESULT 13
Q8H0G3
ID   Q8H0G3      PRELIMINARY;      PRT;   302 AA.
AC   Q8H0G3;
DT   01-MAR-2003 (TrEMBLrel. 23, Created)
DT   01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DT   01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE   Prenyl diphosphate phosphatase.
GN   ATPAP1.
OS   Arabidopsis thaliana (Mouse-ear cress).
```

```
OC   Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC   Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC   eurosids II; Brassicales; Brassicaceae; Arabidopsis.
OX   NCBI_TaxID=3702;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=cv. Columbia;
RA   Tokuhiro K., Muramoto N., Yamada Y., Asami O., Hirai M., Obata S.,
RA   Ohto C., Muramatsu M.;
RT   "Prenyl alcohol production by overexpression of prenyl diphosphate
RT   phosphatase in Yeast Saccharomyces cerevisiae.";
RL   Submitted (JAN-2001) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AB053949; BAC41334.1; -.
SQ   SEQUENCE   302 AA;  33739 MW;  9E1C6D7DAFD569D6 CRC64;

  Query Match            34.8%;  Score 337;  DB 10;  Length 302;
  Best Local Similarity  42.1%;  Pred. No. 5.9e-29;
  Matches   69;  Conservative   31;  Mismatches   54;  Indels   10;  Gaps    4;

Qy         16 TKPMF---VIAFLSPLSLIFLAKFLKKADTRDSRQACLAASLALALNGVFTNTIKLIVGR 72
              | |::   | | | |: ::|:  :||:    |   : |    |: : || |::||:  ||
Db         65 TVPIWSVPVYAVLLPI-IVFVCFYLKRTCVYDLHHSILGLLFAVLITGVITDSIKVATGR 123

Qy         73 PRPDFFYRCFPDGL----AHSDLMCTGDKDVVNEGRKSFPSGHSSFAFAGLAFASFYLAG 128
              |||:|::||||||    |   ::| |    | || |||||||:|::|||| | | ||:|
Db        124 PRPNFYWRCFPDGKELYDALGGVVCHGKAAEVKEGHKSFPSGHTSWSFAGLTFLSLYLSG 183

Qy        129 KLHCFTPQGRGKSWRFCAFLSPLLFAAVIALSRTCDYKHHWQDL 172
              |:  |  :|     : |  : ||| | :: :||  || |||||:
Db        184 KIKAFNNEGHVA--KLCLVIFPLLAACLVGISRVDDYWHHWQDV 225


RESULT 14
Q8LDP8
ID   Q8LDP8      PRELIMINARY;      PRT;   327 AA.
AC   Q8LDP8;
DT   01-OCT-2002 (TrEMBLrel. 22, Created)
DT   01-OCT-2002 (TrEMBLrel. 22, Last sequence update)
DT   01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE   Putative phosphatidic acid phosphatase.
OS   Arabidopsis thaliana (Mouse-ear cress).
OC   Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC   Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC   eurosids II; Brassicales; Brassicaceae; Arabidopsis.
OX   NCBI_TaxID=3702;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Haas B.J., Volfovsky N., Town C.D., Troukhan M., Alexandrov N.,
RA   Feldmann K.A., Flavell R.B., White O., Salzberg S.L.;
RT   "Full-length messenger RNA sequences greatly improve genome
RT   annotation.";
RL   Genome Biol. 0:0-0(2002).
RN   [2]
RP   SEQUENCE FROM N.A.
RA   Brover V., Troukhan M., Alexandrov N., Lu Y.-P., Flavell R.,
RA   Feldmann K.;
RT   "Full-Length cDNA from Arabidopsis thaliana.";
RL   Submitted (MAR-2002) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AY085869; AAM63082.1; -.
DR   InterPro; IPR000326; PA_PTPase.
DR   Pfam; PF01569; PAP2; 1.
DR   SMART; SM00014; acidPPc; 1.
SQ   SEQUENCE   327 AA;  36646 MW;  748B2A02EC3EEA9E CRC64;

  Query Match            34.8%;  Score 337;  DB 10;  Length 327;
  Best Local Similarity  42.1%;  Pred. No. 6.4e-29;
  Matches   69;  Conservative   31;  Mismatches   54;  Indels   10;  Gaps    4;

Qy         16 TKPMF---VIAFLSPLSLIFLAKFLKKADTRDSRQACLAASLALALNGVFTNTIKLIVGR 72
              | |::   | | | |: ::|:  :||:    |   : |    |: : || |::||:  ||
Db         90 TVPIWSVPVYAVLVPI-IVFVCFYLKRTCVYDLHHSILGLLFAVLITGVITDSIKVATGR 148

Qy         73 PRPDFFYRCFPDGL----AHSDLMCTGDKDVVNEGRKSFPSGHSSFAFAGLAFASFYLAG 128
              |||:|::||||||    |   ::| |    | || |||||||:|::|||| | | ||:|
Db        149 PRPNFYWRCFPDGKELYDALGGVVCHGKAAEVKEGHKSFPSGHTSWSFAGLTFLSLYLSG 208

Qy        129 KLHCFTPQGRGKSWRFCAFLSPLLFAAVIALSRTCDYKHHWQDL 172
              |:  |  :|     : |  : ||| | :: :||  || |||||:
Db        209 KIKAFNNEGHVA--KLCLVIFPLLAACLVGISRVDDYWHHWQDV 250


RESULT 15
Q9ZU49
ID   Q9ZU49      PRELIMINARY;      PRT;   327 AA.
AC   Q9ZU49; Q94LY6;
DT   01-MAY-1999 (TrEMBLrel. 10, Created)
DT   01-JUN-2002 (TrEMBLrel. 21, Last sequence update)
DT   01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
DE   Putative phosphatidic acid phosphatase.
GN   AT2G01180 OR ATPAP1.
OS   Arabidopsis thaliana (Mouse-ear cress).
OC   Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC   Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC   eurosids II; Brassicales; Brassicaceae; Arabidopsis.
OX   NCBI_TaxID=3702;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=cv. Columbia;
RA   Lin X., Kaul S., Shea T.P., Fujii C.Y., Shen M., VanAken S.E.,
RA   Barnstead M.E., Mason T.M., Bowman C.L., Ronning C.M., Benito M.-I.,
RA   Carrera A.J., Creasy T.H., Buell C.R., Town C.D., Nierman W.C.,
RA   Fraser C.M., Venter J.C.;
RL   Submitted (FEB-2002) to the EMBL/GenBank/DDBJ databases.
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=Columbia;
RA   Katagiri T.; Shinozaki K.;
RT   "One of the biological functions of the gene encoding phosphatidic
RT   acid phosphatase (PAP) correlates cell elongation in Arabidopsis
RT   thaliana.";
RL   Submitted (MAY-1997) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AC006200; AAD14518.2; -.
DR   EMBL; AB004318; BAB47575.1; -.
DR   InterPro; IPR000326; PA_PTPase.
DR   Pfam; PF01569; PAP2; 1.
DR   SMART; SM00014; acidPPc; 1.
SQ   SEQUENCE   327 AA;  36683 MW;  BAF6F345817BA245 CRC64;

  Query Match            34.8%;  Score 337;  DB 10;  Length 327;
  Best Local Similarity  42.1%;  Pred. No. 6.4e-29;
  Matches   69;  Conservative   31;  Mismatches   54;  Indels   10;  Gaps    4;

Qy         16 TKPMF---VIAFLSPLSLIFLAKFLKKADTRDSRQACLAASLALALNGVFTNTIKLIVGR 72
              | |::   | | | |: ::|:  :||:    |   : |    |: : || |::||:  ||
Db         90 TVPIWSVPVYAVLLPI-IVFVCFYLKRTCVYDLHHSILGLLFAVLITGVITDSIKVATGR 148

Qy         73 PRPDFFYRCFPDGL----AHSDLMCTGDKDVVNEGRKSFPSGHSSFAFAGLAFASFYLAG 128
              |||:|::||||||    |   ::| |    | || |||||||:|::|||| | | ||:|
Db        149 PRPNFYWRCFPDGKELYDALGGVVCHGKAAEVKEGHKSFPSGHTSWSFAGLTFLSLYLSG 208

Qy        129 KLHCFTPQGRGKSWRFCAFLSPLLFAAVIALSRTCDYKHHWQDL 172
              |:  |  :|     : |  : ||| | :: :||  || |||||:
Db        209 KIKAFNNEGHVA--KLCLVIFPLLAACLVGISRVDDYWHHWQDV 250


Search completed: August 14, 2003, 13:07:23
Job time : 96 secs
```

GenCore version 5.1.6

OM protein - protein search, using sw model

Run on: August 14, 2003, 13:03:36 ; Search time 85 Seconds
(without alignments)
336.127 Million cell updates/sec

Title: US-09-876-997-399
Perfect score: 968
Sequence: 1 MWLYRNPYVEAEYFPTKPMF.....RTCDYKHHWQDLLKCTNTAK 180

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 1107863 seqs, 158726573 residues

Total number of hits satisfying chosen parameters: 1107863

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : A_Geneseq_19Jun03:*
1: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1980.DAT:*
2: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1981.DAT:*
3: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1982.DAT:*
4: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1983.DAT:*
5: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1984.DAT:*
6: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1985.DAT:*
7: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1986.DAT:*
8: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1987.DAT:*
9: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1988.DAT:*
10: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1989.DAT:*
11: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1990.DAT:*
12: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1991.DAT:*
13: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1992.DAT:*
14: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1993.DAT:*
15: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1994.DAT:*
16: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1995.DAT:*
17: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1996.DAT:*
18: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1997.DAT:*
19: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1998.DAT:*
20: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1999.DAT:*
21: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA2000.DAT:*
22: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA2001.DAT:*
23: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA2002.DAT:*
24: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA2003.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	968	100.0	180	22	AAG89279	Human secreted pro
2	916	94.6	175	22	AAB70690	Human hDPP protein
3	809	83.6	157	22	AAM23571	Arabidopsis EST en
4	641	66.2	233	22	AAU23516	Novel human enzyme
5	641	66.2	233	22	AAU21835	Novel human neopla
6	641	66.2	271	23	ABB77736	Amino acid sequenc
7	548	56.6	184	22	ABB11177	Human phosphatidat
8	440	45.5	363	22	ABB59271	Drosophila melanog
9	440	45.5	412	22	ABB59272	Drosophila melanog
10	389	40.2	128	21	AAY76061	Rat skin cell tran
11	389	40.2	128	22	AAB56000	Skin cell protein,
12	389	40.2	128	23	ABB72200	Rat protein isolat
13	382	39.5	292	21	AAG05044	Arabidopsis thalia
14	382	39.5	292	21	AAG13844	Arabidopsis thalia
15	382	39.5	292	21	AAG14938	Arabidopsis thalia
16	382	39.5	292	21	AAG17708	Arabidopsis thalia
17	382	39.5	299	21	AAG05043	Arabidopsis thalia
18	382	39.5	299	21	AAG13502	Arabidopsis thalia
19	382	39.5	299	21	AAG13843	Arabidopsis thalia
20	382	39.5	299	21	AAG14937	Arabidopsis thalia
21	382	39.5	299	21	AAG17707	Arabidopsis thalia
22	382	39.5	299	21	AAG22001	Arabidopsis thalia
23	382	39.5	314	21	AAG05042	Arabidopsis thalia
24	382	39.5	314	21	AAG13501	Arabidopsis thalia
25	382	39.5	314	21	AAG13842	Arabidopsis thalia
26	382	39.5	314	21	AAG14936	Arabidopsis thalia
27	382	39.5	314	21	AAG17706	Arabidopsis thalia
28	382	39.5	314	21	AAG22000	Arabidopsis thalia
29	382	39.5	314	21	AAY58789	Arabidopsis phosph
30	382	39.5	314	24	ABU08360	Arabidopsis thalia
31	382	39.5	333	21	AAG13500	Arabidopsis thalia
32	382	39.5	364	21	AAG21999	Arabidopsis thalia
33	377	38.9	322	21	AAY58792	Soybean phosphatid
34	377	38.9	322	24	ABU08362	Soybean phosphatid
35	372	38.4	238	21	AAG30402	Arabidopsis thalia
36	372	38.4	268	21	AAG30401	Arabidopsis thalia
37	372	38.4	290	21	AAG30400	Arabidopsis thalia
38	372	38.4	290	21	AAY58787	Arabidopsis phosph
39	372	38.4	290	23	ABB99108	A. thaliana phosph
40	372	38.4	290	24	ABU08358	Arabidopsis thalia
41	371	38.3	280	21	AAG13025	Arabidopsis thalia
42	371	38.3	287	21	AAG13024	Arabidopsis thalia
43	371	38.3	302	21	AAG13023	Arabidopsis thalia
44	360.5	37.2	115	23	ABB89831	Human polypeptide
45	357	36.9	343	21	AAY58791	Soybean phosphatid

ALIGNMENTS

RESULT 1
AAG89279
ID AAG89279 standard; Protein; 180 AA.
XX
AC AAG89279;
XX
DT 11-SEP-2001 (first entry)
XX
DE Human secreted protein, SEQ ID NO: 399.
XX
KW Human; secreted protein; gene therapy; vaccine; treatment; diagnosis;
KW GENSET.
XX
OS Homo sapiens.
XX
PN WO200142451-A2.
XX
PD 14-JUN-2001.
XX
PF 07-DEC-2000; 2000WO-IB01938.
XX
PR 08-DEC-1999; 99US-0169629.
PR 06-MAR-2000; 2000US-0187470.
XX
PA (GEST) GENSET.
XX
PI Dumas Milne Edwards J, Bougueleret L, Jobert S;
XX
DR WPI; 2001-367870/38.
DR N-PSDB; AAH64882.
XX
PT Full length GENSET human nucleic acids encoding potentially secreted

PT proteins, useful in gene therapy and vaccination against a variety of
PT diseases, and for diagnosis of those diseases -
XX
PS Claim 21; Page 882-883; 921pp; English.
XX
CC The invention relates to full length GENSET human nucleic acids encoding
CC potentially secreted proteins. The nucleic acids and the polypeptides
CC they encode may be used in the prevention, treatment and diagnosis of
CC diseases associated with inappropriate GENSET gene expression. For
CC example, they be used to treat disorders associated with decreased
CC GENSET gene expression by rectifying mutations or deletions in a
CC patient's genome that affect the activity of GENSET or by supplementing
CC the patients own production of GENSET polypeptides. Conversely,
CC antisense nucleic acid molecules may be administered to down regulate
CC GENSET expression by binding with the cells' own genes and preventing
CC their expression. The sense and antisense nucleic acids may also be
CC used as DNA probes in diagnostic assays to detect and quantitate the
CC presence of similar nucleic acid sequences in samples, and hence to
CC determine which patients may be in need of restorative therapy.
CC The GENSET polypeptides may be used as antigens in the production of
CC antibodies and in assays to identify modulators (agonists and
CC antagonists) of GENSET polypeptide expression and activity. The
CC present sequence is a GENSET polypeptide of the invention.
XX
SQ Sequence 180 AA;

Query Match 100.0%; Score 968; DB 22; Length 180;
Best Local Similarity 100.0%; Pred. No. 3.1e-110;
Matches 180; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy          1 MWLYRNPYVEAEYFPTKPMFVIAFLSPLSLIFLAKFLKKADTRDSRQACLAASLALALNG 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MWLYRNPYVEAEYFPTKPMFVIAFLSPLSLIFLAKFLKKADTRDSRQACLAASLALALNG 60

Qy         61 VFTNTIKLIVGRPRPDFFYRCFPDGLAHSDLMCTGDKDVVNEGRKSFPSGHSSFAFAGLA 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 VFTNTIKLIVGRPRPDFFYRCFPDGLAHSDLMCTGDKDVVNEGRKSFPSGHSSFAFAGLA 120

Qy        121 FASFYLAGKLHCFTPQGRGKSWRFCAFLSPLLFAAVIALSRTCDYKHHWQDLLKCTNTAK 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 FASFYLAGKLHCFTPQGRGKSWRFCAFLSPLLFAAVIALSRTCDYKHHWQDLLKCTNTAK 180
```

RESULT 2
AAB70690
ID AAB70690 standard; Protein; 175 AA.
XX
AC AAB70690;
XX
DT 17-MAY-2001 (first entry)
XX
DE Human hDPP protein sequence SEQ ID NO:7.
XX
KW Human; hDPP; diacylglycerol pyrophosphate phosphatase; DPP; detection.
XX
OS Homo sapiens.
XX
PN CN1271009-A.
XX
PD 25-OCT-2000.
XX
PF 17-MAR-2000; 2000CN-0114952.
XX
PR 17-MAR-2000; 2000CN-0114952.
XX
PA (SREN-) SOUTHERN RES CENT NAT HUMAN GENE GROUP.
XX
PI Li N, Xiao H, Liu F;
XX
DR WPI; 2001-081384/10.
DR N-PSDB; AAF74766.
XX
PT New human diacyl glyceropyrophosphate phosphatase protein and its code
PT sequence -
XX
PS Claim 4; Page 17; 19pp; Chinese.
XX
CC The present invention describes a human diacylglycerol pyrophosphate
CC phosphatase (DPP) designated hDPP. hDPP is expressed in normal tissue
CC near cancerous liver cells of a human body. Also described are methods
CC for the preparation and detection of hDPP nucleotide and protein
CC sequences. The present sequence represents human hDPP, as given in the
CC present invention.
XX
SQ Sequence 175 AA;

Query Match 94.6%; Score 916; DB 22; Length 175;
Best Local Similarity 98.3%; Pred. No. 7.1e-104;
Matches 171; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

```
Qy          1 MWLYRNPYVEAEYFPTKPMFVIAFLSPLSLIFLAKFLKKADTRDSRQACLAASLALALNG 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MWLYRNPYVEAEYFPTKPMFVIAFLSPLSLIFLAKFLKKADTRDSRQACLAASLALALNG 60

Qy         61 VFTNTIKLIVGRPRPDFFYRCFPDGLAHSDLMCTGDKDVVNEGRKSFPSGHSSFAFAGLA 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 VFTNTIKLIVGRPRPDFFYRCFPDGLAHSDLMCTGDKDVVNEGRKSFPSGHSSFAFAGLA 120

Qy        121 FASFYLAGKLHCFTPQGRGKSWRFCAFLSPLLFAAVIALSRTCDYKHHWQDLLK 174
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||  |
Db        121 FASFYLAGKLHCFTPQGRGKSWRFCAFLSPLLFAAVIALSRTCDYKHHWQGPFK 174
```

RESULT 3
AAM23571
ID AAM23571 standard; Protein; 157 AA.
XX
AC AAM23571;
XX
DT 12-OCT-2001 (first entry)
XX
DE Arabidopsis EST encoded protein SEQ ID NO: 1096.
XX
KW Human; sheep; pig; cow; fruit fly; yeast; hamster; macaque; horse;
KW tomato; monkey; dog; sea urchin; expressed sequence tag; EST;
KW diagnostics; forensic test; gene mapping; genetic disorder;
KW biodiversity; gene therapy; nutrition.
XX
OS Arabidopsis thaliana.
XX
PN WO200154477-A2.
XX
PD 02-AUG-2001.
XX
PF 25-JAN-2001; 2001WO-US02687.
XX
PR 25-JAN-2000; 2000US-0491404.
PR 17-JUL-2000; 2000US-0617746.
PR 03-AUG-2000; 2000US-0631451.
PR 15-SEP-2000; 2000US-0663870.
XX
PA (HYSE-) HYSEQ INC.
XX
PI Tang YT, Liu C, Zhou P, Qian XB, Wang Z, Chen R, Asundi V;
PI Cao Y, Drmanac RA, Zhang J, Werhman T;
XX
DR WPI; 2001-476164/51.
DR N-PSDB; AAH98230.
XX
PT Isolated polypeptide for treatment of diseases, diagnostics, raising
PT antibodies and research use -
XX
PS Claim 20; Page 821-822; 1275pp; English.
XX
CC The present invention provides the protein and coding sequences of novel
CC proteins from a variety of organisms, including human, dog, cat, horse,

```
CC   cow, pig, hamster, monkey, macaque, yeast, bacteria, fruit fly, sea
CC   urchin and tomato. These were derived from expressed sequence tags (ESTs)
CC   from the organism of interest. They can be used in diagnostics,
CC   forensics, gene mapping, identification of mutations, to assess
CC   biodiversity and for nutritional purposes. The present sequence is a
CC   protein of the invention.
XX
SQ   Sequence   157 AA;

  Query Match             83.6%;  Score 809;  DB 22;  Length 157;
  Best Local Similarity   98.1%;  Pred. No. 8.3e-91;
  Matches  153;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;

Qy         19 MFVIAFLSPLSLIFLAKFLKKADTRDSRQACLAASLALALNGVFTNTIKLIVGRPRPDFF 78
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MFVIAFLSPLSLIFLAKFLKKADTRDSRQACLAASLALALNGVFTNTIKLIVGRPRPDFF 60

Qy         79 YRCFPDGLAHSDLMCTGDKDVVNEGRKSFPSGHSSFAFAGLAFASFYLAGKLHCFTPQGR 138
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 YRCFPDGLAHSDLMCTGDKDVVNEGRKSFPSGHSSFAFAGLAFASFYLAGKLHCFTPQGR 120

Qy        139 GKSWRFCAFLSPLLFAAVIALSRTCDYKHHWQDLLK 174
              ||||||||||||||||||||||||||||||    |
Db        121 GKSWRFCAFLSPLLFAAVIALSRTCDYKHHWQGPFK 156


RESULT 4
AAU23516
ID   AAU23516 standard; Protein; 233 AA.
XX
AC   AAU23516;
XX
DT   18-DEC-2001  (first entry)
XX
DE   Novel human enzyme polypeptide #602.
XX
KW   Human; oxidoreductase enzyme; transferase; hydrolase; lyase; isomerase;
KW   ligase; hyperproliferative disorder; immunodeficiency disorder;
KW   autoimmune disorder; neurological disorder; metabolic disorder;
KW   inflammatory disorder; cardiovascular disorder; reproductive disorder;
KW   blood-related disorder; infectious disorder; cytostatic; anti arthritic;
KW   nephrotropic; anticoagulant.
XX
OS   Homo sapiens.
XX
PN   WO200155301-A2.
XX
PD   02-AUG-2001.
XX
PF   17-JAN-2001; 2001WO-US01239.
XX
PR   31-JAN-2000; 2000US-0179065.
PR   04-FEB-2000; 2000US-0180628.
PR   24-FEB-2000; 2000US-0184664.
PR   02-MAR-2000; 2000US-0186350.
PR   16-MAR-2000; 2000US-0189874.
PR   17-MAR-2000; 2000US-0190076.
PR   18-APR-2000; 2000US-0198123.
PR   19-MAY-2000; 2000US-0205515.
PR   07-JUN-2000; 2000US-0209467.
PR   28-JUN-2000; 2000US-0214886.
PR   30-JUN-2000; 2000US-0215135.
PR   07-JUL-2000; 2000US-0216647.
PR   07-JUL-2000; 2000US-0216880.
PR   11-JUL-2000; 2000US-0217487.
PR   11-JUL-2000; 2000US-0217496.
PR   14-JUL-2000; 2000US-0218290.
PR   26-JUL-2000; 2000US-0220963.
PR   26-JUL-2000; 2000US-0220964.
PR   14-AUG-2000; 2000US-0224518.
PR   14-AUG-2000; 2000US-0224519.
PR   14-AUG-2000; 2000US-0225213.
PR   14-AUG-2000; 2000US-0225214.
PR   14-AUG-2000; 2000US-0225266.
PR   14-AUG-2000; 2000US-0225267.
PR   14-AUG-2000; 2000US-0225268.
PR   14-AUG-2000; 2000US-0225270.
PR   14-AUG-2000; 2000US-0225447.
PR   14-AUG-2000; 2000US-0225757.
PR   14-AUG-2000; 2000US-0225758.
PR   14-AUG-2000; 2000US-0225759.
PR   18-AUG-2000; 2000US-0226279.
PR   22-AUG-2000; 2000US-0226681.
PR   22-AUG-2000; 2000US-0226868.
PR   22-AUG-2000; 2000US-0227182.
PR   23-AUG-2000; 2000US-0227009.
PR   30-AUG-2000; 2000US-0228924.
PR   01-SEP-2000; 2000US-0229287.
PR   01-SEP-2000; 2000US-0229343.
PR   01-SEP-2000; 2000US-0229344.
PR   01-SEP-2000; 2000US-0229345.
PR   05-SEP-2000; 2000US-0229509.
PR   05-SEP-2000; 2000US-0229513.
PR   06-SEP-2000; 2000US-0230437.
PR   06-SEP-2000; 2000US-0230438.
PR   08-SEP-2000; 2000US-0231242.
PR   08-SEP-2000; 2000US-0231243.
PR   08-SEP-2000; 2000US-0231244.
PR   08-SEP-2000; 2000US-0231413.
PR   08-SEP-2000; 2000US-0231414.
PR   08-SEP-2000; 2000US-0232080.
PR   08-SEP-2000; 2000US-0232081.
PR   12-SEP-2000; 2000US-0231968.
PR   14-SEP-2000; 2000US-0232397.
PR   14-SEP-2000; 2000US-0232398.
PR   14-SEP-2000; 2000US-0232399.
PR   14-SEP-2000; 2000US-0232400.
PR   14-SEP-2000; 2000US-0232401.
PR   14-SEP-2000; 2000US-0233063.
PR   14-SEP-2000; 2000US-0233064.
PR   14-SEP-2000; 2000US-0233065.
PR   21-SEP-2000; 2000US-0234223.
PR   21-SEP-2000; 2000US-0234274.
PR   25-SEP-2000; 2000US-0234997.
PR   25-SEP-2000; 2000US-0234998.
PR   26-SEP-2000; 2000US-0235484.
PR   27-SEP-2000; 2000US-0235834.
PR   27-SEP-2000; 2000US-0235836.
PR   29-SEP-2000; 2000US-0236327.
PR   29-SEP-2000; 2000US-0236367.
PR   29-SEP-2000; 2000US-0236368.
PR   29-SEP-2000; 2000US-0236369.
PR   29-SEP-2000; 2000US-0236370.
PR   02-OCT-2000; 2000US-0236802.
PR   02-OCT-2000; 2000US-0237037.
PR   02-OCT-2000; 2000US-0237038.
PR   02-OCT-2000; 2000US-0237039.
PR   02-OCT-2000; 2000US-0237040.
PR   13-OCT-2000; 2000US-0239935.
PR   13-OCT-2000; 2000US-0239937.
PR   20-OCT-2000; 2000US-0240960.
PR   20-OCT-2000; 2000US-0241221.
PR   20-OCT-2000; 2000US-0241785.
PR   20-OCT-2000; 2000US-0241786.
PR   20-OCT-2000; 2000US-0241787.
PR   20-OCT-2000; 2000US-0241808.
PR   20-OCT-2000; 2000US-0241809.
PR   20-OCT-2000; 2000US-0241826.
PR   01-NOV-2000; 2000US-0244617.
PR   08-NOV-2000; 2000US-0246474.
PR   08-NOV-2000; 2000US-0246475.
PR   08-NOV-2000; 2000US-0246476.
PR   08-NOV-2000; 2000US-0246477.
PR   08-NOV-2000; 2000US-0246478.
PR   08-NOV-2000; 2000US-0246523.
PR   08-NOV-2000; 2000US-0246524.
```

PR 08-NOV-2000; 2000US-0246525.
PR 08-NOV-2000; 2000US-0246526.
PR 08-NOV-2000; 2000US-0246527.
PR 08-NOV-2000; 2000US-0246528.
PR 08-NOV-2000; 2000US-0246532.
PR 08-NOV-2000; 2000US-0246609.
PR 08-NOV-2000; 2000US-0246610.
PR 08-NOV-2000; 2000US-0246611.
PR 08-NOV-2000; 2000US-0246613.
PR 17-NOV-2000; 2000US-0249207.
PR 17-NOV-2000; 2000US-0249208.
PR 17-NOV-2000; 2000US-0249209.
PR 17-NOV-2000; 2000US-0249210.
PR 17-NOV-2000; 2000US-0249211.
PR 17-NOV-2000; 2000US-0249212.
PR 17-NOV-2000; 2000US-0249213.
PR 17-NOV-2000; 2000US-0249214.
PR 17-NOV-2000; 2000US-0249215.
PR 17-NOV-2000; 2000US-0249216.
PR 17-NOV-2000; 2000US-0249217.
PR 17-NOV-2000; 2000US-0249218.
PR 17-NOV-2000; 2000US-0249244.
PR 17-NOV-2000; 2000US-0249245.
PR 17-NOV-2000; 2000US-0249264.
PR 17-NOV-2000; 2000US-0249265.
PR 17-NOV-2000; 2000US-0249297.
PR 17-NOV-2000; 2000US-0249299.
PR 17-NOV-2000; 2000US-0249300.
PR 01-DEC-2000; 2000US-0250160.
PR 01-DEC-2000; 2000US-0250391.
PR 05-DEC-2000; 2000US-0251030.
PR 05-DEC-2000; 2000US-0251988.
PR 05-DEC-2000; 2000US-0256719.
PR 06-DEC-2000; 2000US-0251479.
PR 08-DEC-2000; 2000US-0251856.
PR 08-DEC-2000; 2000US-0251868.
PR 08-DEC-2000; 2000US-0251869.
PR 08-DEC-2000; 2000US-0251989.
PR 08-DEC-2000; 2000US-0251990.
PR 11-DEC-2000; 2000US-0254097.
PR 05-JAN-2001; 2001US-0259678.
XX
PA (HUMA-) HUMAN GENOME SCI INC.
XX
PI Rosen CA, Barash SC, Ruben SM;
XX
DR WPI; 2001-465566/50.
DR N-PSDB; AAS41386.
XX
PT Novel polypeptides and polynucleotides useful for diagnosing,
PT preventing, treating neural, immune system, muscular, reproductive,
PT pulmonary, cardiovascular, renal, proliferative disorders and cancerous
PT diseases -
XX
PS Claim 11; SEQ ID No 1512; 1180pp; English.
XX
CC The present invention relates to the isolation of novel human enzyme
CC polypeptides, and the cDNA (AAS40785-AAS41684) and genomic sequences
CC encoding them. The enzyme polypeptides of the invention may comprise the
CC functional classes of oxidoreductases, transferases, hydrolases, lyases,
CC isomerases or ligases. The sequences of the invention are useful in the
CC diagnosis, treatment, prevention and/or prognosis of a wide range of
CC disorders including hyperproliferative disorders (e.g. cancer),
CC immunodeficiency disorders (e.g. AIDS) autoimmune disorders
CC (e.g. arthritis), neurological disorders (e.g. Alzheimer's disease),
CC metabolic disorders (e.g. phenylketonuria), inflammatory disorders
CC (e.g. asthma), cardiovascular disorders (e.g. atherosclerosis),
CC blood-related disorders (e.g. haemophilia), reproductive disorders
CC (e.g. infertility) and infectious disorders (e.g. Influenza). The
CC polynucleotides of the invention can also be used in gene therapy.
CC AAU22915-AAU23814 represent the novel human enzyme polypeptides of the
CC invention.
CC Note: The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 233 AA;

Query Match 66.2%; Score 641; DB 22; Length 233;
Best Local Similarity 66.7%; Pred. No. 5.7e-70;
Matches 114; Conservative 27; Mismatches 30; Indels 0; Gaps 0;

```
Qy          1 MWLYRNPYVEAEYFPTKPMFVIAFLSPLSLIFLAKFLKKADTRDSRQACLAASLALALNG 60
              :|||:|| |:::  ||: || |:||:||::| : | ::: |  : ::| || |||||||
Db         47 IWLYKNPLVQSDNIPTRLMFAISFLTPLAVICVVKIIRRTDKTEIKEAFLAVSLALALNG 106

Qy         61 VFTNTIKLIVGRPRPDFFYRCFPDGLAHSDLMCTGDKDVVNEGRKSFPSGHSSFAFAGLA 120
              | ||||||||||||||||||||||: :|:: |||| |:|:|||||||| ||||||:||
Db        107 VCTNTIKLIVGRPRPDFFYRCFPDGVMNSEMHCTGDPDLVSEGRKSFPSIHSSFAFSGLG 166

Qy        121 FASFYLAGKLHCFTPQGRGKSWRFCAFLSPLLFAAVIALSRTCDYKHHWQD 171
              | :|||||||||||  ||||||| || : ||  | :||||| |||||||||
Db        167 FTTFYLAGKLHCFTESGRGKSWRLCAAILPLYCAMMIALSRMCDYKHHWQD 217
```

RESULT 5
AAU21835
ID AAU21835 standard; Protein; 233 AA.
XX
AC AAU21835;
XX
DT 06-DEC-2001 (first entry)
XX
DE Novel human neoplastic disease associated polypeptide #268.
XX
KW Human; neoplastic disease associated polypeptide; cancer;
KW hyperproliferative disorder; neural disorder; immune system disorder;
KW muscular disorder; reproductive disorder; gastrointestinal disorder;
KW pulmonary disorder; cardiovascular disorder; renal disorder;
KW neuroprotective; cytostatic; anti inflammatory; vasotropic.
XX
OS Homo sapiens.
XX
PN WO200155163-A1.
XX
PD 02-AUG-2001.
XX
PF 17-JAN-2001; 2001WO-US01358.
XX
PR 31-JAN-2000; 2000US-0179065.
PR 04-FEB-2000; 2000US-0180628.
PR 24-FEB-2000; 2000US-0184664.
PR 02-MAR-2000; 2000US-0186350.
PR 16-MAR-2000; 2000US-0189874.
PR 17-MAR-2000; 2000US-0190076.
PR 18-APR-2000; 2000US-0198123.
PR 19-MAY-2000; 2000US-0205515.
PR 07-JUN-2000; 2000US-0209467.
PR 28-JUN-2000; 2000US-0214886.
PR 30-JUN-2000; 2000US-0215135.
PR 07-JUL-2000; 2000US-0216647.
PR 07-JUL-2000; 2000US-0216880.
PR 11-JUL-2000; 2000US-0217487.
PR 11-JUL-2000; 2000US-0217496.
PR 14-JUL-2000; 2000US-0218290.
PR 26-JUL-2000; 2000US-0220963.
PR 26-JUL-2000; 2000US-0220964.
PR 14-AUG-2000; 2000US-0224518.
PR 14-AUG-2000; 2000US-0224519.
PR 14-AUG-2000; 2000US-0225213.
PR 14-AUG-2000; 2000US-0225214.
PR 14-AUG-2000; 2000US-0225266.
PR 14-AUG-2000; 2000US-0225267.
PR 14-AUG-2000; 2000US-0225268.
PR 14-AUG-2000; 2000US-0225270.
PR 14-AUG-2000; 2000US-0225447.

PR 14-AUG-2000; 2000US-0225757.
PR 14-AUG-2000; 2000US-0225758.
PR 14-AUG-2000; 2000US-0225759.
PR 18-AUG-2000; 2000US-0226279.
PR 22-AUG-2000; 2000US-0226681.
PR 22-AUG-2000; 2000US-0226868.
PR 22-AUG-2000; 2000US-0227182.
PR 23-AUG-2000; 2000US-0227009.
PR 30-AUG-2000; 2000US-0228924.
PR 01-SEP-2000; 2000US-0229287.
PR 01-SEP-2000; 2000US-0229343.
PR 01-SEP-2000; 2000US-0229344.
PR 01-SEP-2000; 2000US-0229345.
PR 05-SEP-2000; 2000US-0229509.
PR 05-SEP-2000; 2000US-0229513.
PR 06-SEP-2000; 2000US-0230437.
PR 06-SEP-2000; 2000US-0230438.
PR 08-SEP-2000; 2000US-0231242.
PR 08-SEP-2000; 2000US-0231243.
PR 08-SEP-2000; 2000US-0231244.
PR 08-SEP-2000; 2000US-0231413.
PR 08-SEP-2000; 2000US-0231414.
PR 08-SEP-2000; 2000US-0232080.
PR 08-SEP-2000; 2000US-0232081.
PR 12-SEP-2000; 2000US-0231968.
PR 14-SEP-2000; 2000US-0232397.
PR 14-SEP-2000; 2000US-0232398.
PR 14-SEP-2000; 2000US-0232399.
PR 14-SEP-2000; 2000US-0232400.
PR 14-SEP-2000; 2000US-0232401.
PR 14-SEP-2000; 2000US-0233063.
PR 14-SEP-2000; 2000US-0233064.
PR 14-SEP-2000; 2000US-0233065.
PR 21-SEP-2000; 2000US-0234223.
PR 21-SEP-2000; 2000US-0234274.
PR 25-SEP-2000; 2000US-0234997.
PR 25-SEP-2000; 2000US-0234998.
PR 26-SEP-2000; 2000US-0235484.
PR 27-SEP-2000; 2000US-0235834.
PR 27-SEP-2000; 2000US-0235836.
PR 29-SEP-2000; 2000US-0236327.
PR 29-SEP-2000; 2000US-0236367.
PR 29-SEP-2000; 2000US-0236368.
PR 29-SEP-2000; 2000US-0236369.
PR 29-SEP-2000; 2000US-0236370.
PR 02-OCT-2000; 2000US-0236802.
PR 02-OCT-2000; 2000US-0237037.
PR 02-OCT-2000; 2000US-0237038.
PR 02-OCT-2000; 2000US-0237039.
PR 02-OCT-2000; 2000US-0237040.
PR 13-OCT-2000; 2000US-0239935.
PR 13-OCT-2000; 2000US-0239937.
PR 20-OCT-2000; 2000US-0240960.
PR 20-OCT-2000; 2000US-0241221.
PR 20-OCT-2000; 2000US-0241785.
PR 20-OCT-2000; 2000US-0241786.
PR 20-OCT-2000; 2000US-0241787.
PR 20-OCT-2000; 2000US-0241808.
PR 20-OCT-2000; 2000US-0241809.
PR 20-OCT-2000; 2000US-0241826.
PR 01-NOV-2000; 2000US-0244617.
PR 08-NOV-2000; 2000US-0246474.
PR 08-NOV-2000; 2000US-0246475.
PR 08-NOV-2000; 2000US-0246476.
PR 08-NOV-2000; 2000US-0246477.
PR 08-NOV-2000; 2000US-0246478.
PR 08-NOV-2000; 2000US-0246523.
PR 08-NOV-2000; 2000US-0246524.
PR 08-NOV-2000; 2000US-0246525.
PR 08-NOV-2000; 2000US-0246526.
PR 08-NOV-2000; 2000US-0246527.
PR 08-NOV-2000; 2000US-0246528.
PR 08-NOV-2000; 2000US-0246532.
PR 08-NOV-2000; 2000US-0246609.
PR 08-NOV-2000; 2000US-0246610.
PR 08-NOV-2000; 2000US-0246611.
PR 08-NOV-2000; 2000US-0246613.
PR 17-NOV-2000; 2000US-0249207.
PR 17-NOV-2000; 2000US-0249208.
PR 17-NOV-2000; 2000US-0249209.
PR 17-NOV-2000; 2000US-0249210.
PR 17-NOV-2000; 2000US-0249211.
PR 17-NOV-2000; 2000US-0249212.
PR 17-NOV-2000; 2000US-0249213.
PR 17-NOV-2000; 2000US-0249214.
PR 17-NOV-2000; 2000US-0249215.
PR 17-NOV-2000; 2000US-0249216.
PR 17-NOV-2000; 2000US-0249217.
PR 17-NOV-2000; 2000US-0249218.
PR 17-NOV-2000; 2000US-0249244.
PR 17-NOV-2000; 2000US-0249245.
PR 17-NOV-2000; 2000US-0249264.
PR 17-NOV-2000; 2000US-0249265.
PR 17-NOV-2000; 2000US-0249297.
PR 17-NOV-2000; 2000US-0249299.
PR 17-NOV-2000; 2000US-0249300.
PR 01-DEC-2000; 2000US-0250160.
PR 01-DEC-2000; 2000US-0250391.
PR 05-DEC-2000; 2000US-0251030.
PR 05-DEC-2000; 2000US-0251988.
PR 05-DEC-2000; 2000US-0256719.
PR 06-DEC-2000; 2000US-0251479.
PR 08-DEC-2000; 2000US-0251856.
PR 08-DEC-2000; 2000US-0251868.
PR 08-DEC-2000; 2000US-0251869.
PR 08-DEC-2000; 2000US-0251989.
PR 08-DEC-2000; 2000US-0251990.
PR 11-DEC-2000; 2000US-0254097.
PR 05-JAN-2001; 2001US-0259678.
XX
PA (HUMA-) HUMAN GENOME SCI INC.
XX
PI Rosen CA, Barash SC, Ruben SM;
XX
DR WPI; 2001-465558/50.
DR N-PSDB; AAS35034.
XX
PT Novel polypeptides and polynucleotides useful as diagnostic reagents to
PT diagnose diseases or disorders associated with aberrant expression or
PT activity of polypeptides, and for treating cancers, rheumatoid
PT arthritis -
XX
PS Claim 11; SEQ ID No 562; 687pp; English.
XX
CC The present invention relates to the isolation of novel human neoplastic
CC disease associated polypeptides, and cDNA (AAS34767-AAS35050) and DNA
CC sequences encoding for these polypeptides. The sequences of the
CC invention are useful in the diagnosis, treatment, prevention and/or
CC prognosis of disorders involving neoplastic disease such as
CC hyperproliferative disorders (e.g. leukaemia, bone cancer, bladder
CC cancer, brain stem glioma, adult liver cancer, childhood cerebellar
CC astrocytoma, or Hodgkin's lymphoma). The sequences of the invention may
CC also be useful for treating other disorders such as neural disorders,
CC immune system disorders, muscular disorders, reproductive disorders,
CC gastrointestinal disorders, pulmonary disorders, cardiovascular disorders
CC and renal disorders. The polynucleotide sequences of the invention are
CC also useful in gene therapy. AAU21568-AAU21851 represent the novel human
CC neoplastic disease associated polypeptides of the invention.
CC Note: The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 233 AA;

Query Match 66.2%; Score 641; DB 22; Length 233;
Best Local Similarity 66.7%; Pred. No. 5.7e-70;

```
  Matches  114;  Conservative   27;  Mismatches   30;  Indels    0;  Gaps    0;

Qy          1 MWLYRNPYVEAEYFPTKPMFVIAFLSPLSLIFLAKFLKKADTRDSRQACLAASLALALNG 60
              :|||:|| |:::  ||: || |:||:||::| : | ::: |  : ::| || ||||||||
Db         47 IWLYKNPLVQSDNIPTRLMFAISFLTPLAVICVVKIIRRTDKTEIKEAFLAVSLALALNG 106

Qy         61 VFTNTIKLIVGRPRPDFFYRCFPDGLAHSDLMCTGDKDVVNEGRKSFPSGHSSFAFAGLA 120
              | ||||||||||||||||||||||: :|:: |||| |:|:|||||||| ||||||:||
Db        107 VCTNTIKLIVGRPRPDFFYRCFPDGVMNSEMHCTGDPDLVSEGRKSFPSIHSSFAFSGLG 166

Qy        121 FASFYLAGKLHCFTPQGRGKSWRFCAFLSPLLFAAVIALSRTCDYKHHWQD 171
              | :|||||||||||  ||||||| || : ||  | :||||| |||||||||
Db        167 FTTFYLAGKLHCFTESGRGKSWRLCAAILPLYCAMMIALSRMCDYKHHWQD 217


RESULT 6
ABB77736
ID   ABB77736 standard; Protein; 271 AA.
XX
AC   ABB77736;
XX
DT   30-JUL-2002  (first entry)
XX
DE   Amino acid sequence of human phosphatidic acid phosphatase 29.81.
XX
KW   Human; phosphatidic acid phosphatase 29.81; tumour; hemopathy; enzyme;
KW   HIV infection; immunological disease; inflammation; gene therapy.
XX
OS   Homo sapiens.
XX
PN   WO200226798-A1.
XX
PD   04-APR-2002.
XX
PF   20-AUG-2001; 2001WO-CN01253.
XX
PR   23-AUG-2000; 2000CN-0119718.
XX
PA   (SHAN-) SHANGHAI BIOWINDOW GENE DEV INC.
PA   (MAOY/) MAO Y.
PA   (XIEY/) XIE Y.
XX
DR   WPI; 2002-292479/33.
DR   N-PSDB; ABL56714.
XX
PT   Phosphatidic acid phosphatase 29.81 and encoding polynucleotide, used
PT   in diagnosis and treatment of malignant tumours, hemopathy, human
PT   immunodeficiency virus infection, immunological diseases and
PT   inflammation -
XX
PS   Claim 1; Page 29; 35pp; Chinese.
XX
CC   The present sequence represents human phosphatidic acid phosphatase
CC   29.81. The 29.81 polynucleotide and polypeptide are used in
CC   diagnosis and treatment of malignant tumour, hemopathy, human
CC   immunodeficiency virus (HIV) infection, immunological diseases and
CC   various inflammations. The polynucleotide may also be used for gene
CC   therapy.
XX
SQ   Sequence   271 AA;

  Query Match             66.2%;  Score 641;  DB 23;  Length 271;
  Best Local Similarity   66.7%;  Pred. No. 7e-70;
  Matches  114;  Conservative   27;  Mismatches   30;  Indels    0;  Gaps    0;

Qy          1 MWLYRNPYVEAEYFPTKPMFVIAFLSPLSLIFLAKFLKKADTRDSRQACLAASLALALNG 60
              :|||:|| |:::  ||: || |:||:||::| : | ::: |  : ::| || ||||||||
Db         36 IWLYKNPLVQSDNIPTRLMFAISFLTPLAVICVVKIIRRTDKTEIKEAFLAVSLALALNG 95

Qy         61 VFTNTIKLIVGRPRPDFFYRCFPDGLAHSDLMCTGDKDVVNEGRKSFPSGHSSFAFAGLA 120
              | ||||||||||||||||||||||: :|:: |||| |:|:|||||||| ||||||:||
Db         96 VCTNTIKLIVGRPRPDFFYRCFPDGVMNSEMHCTGDPDLVSEGRKSFPSIHSSFAFSGLG 155

Qy        121 FASFYLAGKLHCFTPQGRGKSWRFCAFLSPLLFAAVIALSRTCDYKHHWQD 171
              | :|||||||||||  ||||||| || : ||  | :||||| |||||||||
Db        156 FTTFYLAGKLHCFTESGRGKSWRLCAAILPLYCAMMIALSRMCDYKHHWQD 206


RESULT 7
ABB11177
ID   ABB11177 standard; peptide; 184 AA.
XX
AC   ABB11177;
XX
DT   11-JAN-2002  (first entry)
XX
DE   Human phosphatidate phosphohydrolase homologue, SEQ ID NO:1547.
XX
KW   Human; cytokine; cell proliferation; cell differentiation; growth factor;
KW   haematopoiesis regulation; tissue growth; immunomodulator; activin;
KW   inhibin; chemotaxis; chemokinesis; thrombolysis; oncogenesis;
KW   proliferation; metastasis; cancer; tumour; haematopoietic disorder;
KW   myeloid cell disorder; lymphoid cell disorder; asthma; arthritis;
KW   chronic inflammatory condition; proliferative retinopathy;
KW   atherosclerosis; coronary heart disease; arterial ischaemia;
KW   bone disorder; osteoporosis; vascular growth disorder;
KW   tissue regeneration; wound healing; infection; immune disorder;
KW   cell culture; drug screening; gene therapy; antiinflammatory;
KW   antiasthmatic; antiarthritic; haemostatic; antiarteriosclerotic;
KW   cytostatic; osteopathic; vasotropic; cardiant; virucide; antibacterial;
KW   antifungal; vulnerary; antiulcer.
XX
OS   Homo sapiens.
XX
PN   WO200157188-A2.
XX
PD   09-AUG-2001.
XX
PF   05-FEB-2001; 2001WO-US03800.
XX
PR   03-FEB-2000; 2000US-0496914.
PR   27-APR-2000; 2000US-0560875.
XX
PA   (HYSE-) HYSEQ INC.
XX
PI   Tang YT,  Liu C,  Drmanac RT;
XX
DR   WPI; 2001-457740/49.
DR   N-PSDB; ABA08421.
XX
PT   Human proteins and DNA encoding sequences useful for preventing,
PT   treating or ameliorating a medical condition in a mammalian subject
PT   e.g. arthritis and cancer -
XX
PS   Claim 20; Page 153; 1963pp; English.
XX
CC   Sequences ABB10981-ABB12330 represent 1350 novel human polypeptides, and
CC   sequences ABA08225-ABA09574 represent nucleic acids encoding them. The
CC   invention also relates to vectors and recombinant host cells comprising a
CC   nucleotide of the invention, methods of producing the novel polypeptides,
CC   antibodies against the polypeptides, methods of detecting the nucleotides
CC   or polypeptides in a sample, and methods of identifying compounds which
CC   bind to polypeptides of the invention. Although novel, many of the
CC   polypeptides of the invention have homology to known proteins, thereby
CC   giving an insight into their probable biological activities, and hence
CC   potential therapeutic applications. The polypeptides of the invention may
CC   have various activities, including cytokine, cell proliferation or cell
CC   differentiation activities; stem cell growth factor activity;
CC   haematopoiesis regulatory activity; tissue growth activity;
CC   immunomodulatory activity; activin- or inhibin-related activities;
CC   chemotactic or chemokinetic activities; haemostatic, thrombotic or
CC   thrombolytic activities; receptor or ligand activities; or may be
CC   involved in oncogenesis, cancer cell proliferation or metastasis.
CC   Depending on their biological activities, polypeptides and nucleotides of
CC   the invention are useful for preventing, treating or ameliorating medical
CC   conditions, e.g., by protein or gene therapy. Such conditions include
```

CC cancers, haematopoietic disorders (e.g.,myeloid or lymphoid cell
CC disorders), chronic inflammatory conditions (e.g., asthma or arthritis),
CC proliferative retinopathy, atherosclerosis, coronary heart disease,
CC arterial ischaemia, bone disorders (e.g., osteoporosis), and abnormal
CC vascular growth. Polypeptides involved with tissue regeneration and
CC repair (or nucleic acids encoding them) may be used to promote wound
CC healing (e.g., of burns, incisions and ulcers), while those with
CC immunomodulatory activities may be used in the treatment of viral,
CC bacterial and fungal infections in addition to immune disorders.
CC Polypeptides with growth factor activity may be used in cell cultures to
CC promote cell growth. For example, such polypeptides may be used to
CC manipulate stem cells in culture to give rise to neuroepithelial cells
CC that can be used to augment or replace cells damaged by illness,
CC autoimmune disease or accidental damage. The polypeptides and nucleotides
CC may also be used in the diagnosis of the above conditions, and in drug
CC screening techniques. The present sequence represents a novel human
CC polypeptide of the invention.
XX
SQ Sequence 184 AA;

Query Match 56.6%; Score 548; DB 22; Length 184;
Best Local Similarity 64.9%; Pred. No. 1.1e-58;
Matches 98; Conservative 26; Mismatches 27; Indels 0; Gaps 0;

```
Qy          1 MWLYRNPYVEAEYFPTKPMFVIAFLSPLSLIFLAKFLKKADTRDSRQACLAASLALALNG 60
              :|||:||  :::  ||: || |:||:||::| : | ::: |  : ::| || ||||||||
Db         34 IWLYKNPLGQSDNIPTRLMFAISFLTPLAVICVVKIIRRTDKTEIKEAFLAVSLALALNG 93

Qy         61 VFTNTIKLIVGRPRPDFFYRCFPDGLAHSDLMCTGDKDVVNEGRKSFPSGHSSFAFAGLA 120
              | ||||||||||||||||||||||||: :|:: |||| |:|:|||||||| ||||||:||
Db         94 VCTNTIKLIVGRPRPDFFYRCFPDGVMNSEMHCTGDPDLVSEGRKSFPSIHSSFAFSGLG 153

Qy        121 FASFYLAGKLHCFTPQGRGKSWRFCAFLSPL 151
              | :|||||||||||  ||||||| || : ||
Db        154 FTTFYLAGKLHCFTESGRGKSWRLCAAILPL 184
```

RESULT 8
ABB59271
ID ABB59271 standard; Protein; 363 AA.
XX
AC ABB59271;
XX
DT 26-MAR-2002 (first entry)
XX
DE Drosophila melanogaster polypeptide SEQ ID NO 4605.
XX
KW Drosophila; developmental biology; cell signalling; insecticide;
KW pharmaceutical.
XX
OS Drosophila melanogaster.
XX
PN WO200171042-A2.
XX
PD 27-SEP-2001.
XX
PF 23-MAR-2001; 2001WO-US09231.
XX
PR 23-MAR-2000; 2000US-191637P.
PR 11-JUL-2000; 2000US-0614150.
XX
PA (PEKE) PE CORP NY.
XX
PI Venter JC, Adams M, Li PWD, Myers EW;
XX
DR WPI; 2001-656860/75.
DR N-PSDB; ABL03374.
XX
PT New isolated nucleic acid detection reagent for detecting 1000 or more
PT genes from Drosophila and for elucidating cell signalling and cell-cell
PT interactions -
XX
PS Disclosure; SEQ ID NO 4605; 21pp + Sequence Listing; English.
XX
CC The invention relates to an isolated nucleic acid detection reagent
CC capable of detecting 1000 or more genes from Drosophila. The invention is
CC useful in developmental biology and in elucidating cell signalling and
CC cell-cell interactions in higher eukaryotes for the development of
CC insecticides, therapeutics and pharmaceutical drugs. The invention
CC discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA
CC sequences (ABL01840-ABL16175) and the encoded proteins
CC (ABB57737-ABB72072).
CC The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 363 AA;

Query Match 45.5%; Score 440; DB 22; Length 363;
Best Local Similarity 44.5%; Pred. No. 4.7e-45;
Matches 85; Conservative 29; Mismatches 49; Indels 28; Gaps 4;

```
Qy          1 MWLYRNPYVEAEYFPTKP-------MFVIAFLSPLSLIFLAKFLKKADTRDSRQACLAAS 53
              :|||:|        | :|       :     ::|  |: :| :    | || | |  | :
Db        103 LWLYKN--------PRRPDIVRGGELLFWVIVAPF-LVTIAFYWYTRDRRDFRAASWAWT 153

Qy         54 LALALNGVFTNTIKLIVGRPRPDFFYRCFPDGL-----------AHSDLMCTGDKDVVN 101
              ||| :||: |: :|: |||||||:||||||||:           :  |  |||    :|
Db        154 LALCMNGIPTSVLKITVGRPRPDYFYRCFPDGVMVLNTTSNGVDTSILDFNCTGLPGDIN 213

Qy        102 EGRKSFPSGHSSFAFAGLAFASFYLAGKLHCFTPQGRGKSWRFCAFLSPLLFAAVIALSR 161
              |||||||||||||||   | ::|:  ||| |  :||| :|| |  : ||  | ::|:||
Db        214 EGRKSFPSGHSSFAFASFGFIAYYIGAKLHAFDSRGRGHTWRLCIAVIPLFIALLVAVSR 273

Qy        162 TCDYKHHWQDL 172
              |||| |||||:
Db        274 TCDYHHHWQDV 284
```

RESULT 9
ABB59272
ID ABB59272 standard; Protein; 412 AA.
XX
AC ABB59272;
XX
DT 26-MAR-2002 (first entry)
XX
DE Drosophila melanogaster polypeptide SEQ ID NO 4608.
XX
KW Drosophila; developmental biology; cell signalling; insecticide;
KW pharmaceutical.
XX
OS Drosophila melanogaster.
XX
PN WO200171042-A2.
XX
PD 27-SEP-2001.
XX
PF 23-MAR-2001; 2001WO-US09231.
XX
PR 23-MAR-2000; 2000US-191637P.
PR 11-JUL-2000; 2000US-0614150.
XX
PA (PEKE) PE CORP NY.
XX
PI Venter JC, Adams M, Li PWD, Myers EW;
XX
DR WPI; 2001-656860/75.
DR N-PSDB; ABL03375.
XX
PT New isolated nucleic acid detection reagent for detecting 1000 or more
PT genes from Drosophila and for elucidating cell signalling and cell-cell
PT interactions -
XX
PS Disclosure; SEQ ID NO 4608; 21pp + Sequence Listing; English.
XX

```
CC   The invention relates to an isolated nucleic acid detection reagent
CC   capable of detecting 1000 or more genes from Drosophila. The invention is
CC   useful in developmental biology and in elucidating cell signalling and
CC   cell-cell interactions in higher eukaryotes for the development of
CC   insecticides, therapeutics and pharmaceutical drugs. The invention
CC   discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA
CC   sequences (ABL01840-ABL16175) and the encoded proteins
CC   (ABB57737-ABB72072).
CC   The sequence data for this patent did not form part of the printed
CC   specification, but was obtained in electronic format directly from WIPO
CC   at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ   Sequence   412 AA;

  Query Match             45.5%;  Score 440;  DB 22;  Length 412;
  Best Local Similarity   44.5%;  Pred. No. 5.5e-45;
  Matches   85;  Conservative   29;  Mismatches   49;  Indels   28;  Gaps    4;

Qy          1 MWLYRNPYVEAEYFPTKP-------MFVIAFLSPLSLIFLAKFLKKADTRDSRQACLAAS 53
              :|||:|        | :|       :     ::|  |: :| :    | || | |  | :
Db        152 LWLYKN--------PRRPDIVRGGELLFWVIVAPF-LVTIAFYWYTRDRRDFRAASWAWT 202

Qy         54 LALALNGVFTNTIKLIVGRPRPDFFYRCFPDGL-----------AHSDLMCTGDKDVVN 101
              ||| :||: |: :|: |||||||:||||||||:           :  |  |||    :|
Db        203 LALCMNGIPTSVLKITVGRPRPDYFYRCFPDGVMVLNTTSNGVDTSILDFNCTGLPGDIN 262

Qy        102 EGRKSFPSGHSSFAFAGLAFASFYLAGKLHCFTPQGRGKSWRFCAFLSPLLFAAVIALSR 161
              |||||||||||||||   | ::|:  ||| |  :||| :|| |  : ||  | ::|:||
Db        263 EGRKSFPSGHSSFAFASFGFIAYYIGAKLHAFDSRGRGHTWRLCIAVIPLFIALLVAVSR 322

Qy        162 TCDYKHHWQDL 172
              |||| |||||:
Db        323 TCDYHHHWQDV 333


RESULT 10
AAY76061
ID   AAY76061 standard; Protein; 128 AA.
XX
AC   AAY76061;
XX
DT   27-MAR-2000  (first entry)
XX
DE   Rat skin cell transmembrane protein, SEQ ID NO:316.
XX
KW   Skin; dermal papilla; keratinocyte; neonatal foreskin fibroblast;
KW   embryonic skin cell; keratinocyte stem cell; transit amplifying cell;
KW   secreted; transmembrane; inflammation; cancer; neurological disease;
KW   angiogenesis; tumour vascularisation; growth disorder;
KW   developmental disorder; skin wound; hair follicle disorder;
KW   anti-inflammatory; cytostatic; neuroprotective; vulnery.
XX
OS   Rattus sp.
XX
PN   WO9955865-A1.
XX
PD   04-NOV-1999.
XX
PF   29-APR-1999;   99WO-NZ00051.
XX
PR   29-APR-1998;   98US-0069726.
PR   09-NOV-1998;   98US-0188930.
XX
PA   (GENE-) GENESIS RES & DEV CORP LTD.
XX
PI   Strachan L,  Sleeman M,  Watson JD,  Onrust R,  Kumble A,  Murison JG;
XX
DR   WPI; 2000-072177/06.
XX
PT   Novel polynucleotides useful for the treatment of various conditions
PT   including wounds and cancer -
XX
PS   Claim 4; Page 186-187; 235pp; English.
XX
CC   The invention relates to novel nucleic acid sequences derived from rat
CC   dermal papilla, human keratinocytes and neonatal foreskin fibroblasts,
CC   and mouse embryonic skin, keratinocyte stem cells and transit amplifying
CC   cells. Polypeptides of the invention may be used to treat inflammation,
CC   cancer and neurological diseases. The proteins may be used to stimulate
CC   the growth and motility of keratinocytes, to inhibit the growth of
CC   cancer cells, to modulate angiogenesis and tumour vascularisation, to
CC   modulate skin inflammation, to modulate epithelial cell growth and to
CC   inhibit binding of HIV-1 to leukocytes. The invention may also be used
CC   to treat growth and developmental defects, skin wounds and hair follicle
CC   disorders. Sequences AAY75942-Y76123 represent polypeptides encoded
CC   by cDNA sequences derived from several mouse, rat or human skin cell
CC   types. Sequences AAY75942-Y75947, AAY76020-Y76021, AAY76094-Y76104 and
CC   AAY76119 are proteins with an N-terminal signal sequence, indicating
CC   that they are secreted. Sequences AAY75986-Y75989, AAY76061-Y76071,
CC   AAY76106-Y76109 and AAY76121-Y76122 are proteins with one or more
CC   putative transmembrane domains.
XX
SQ   Sequence   128 AA;

  Query Match             40.2%;  Score 389;  DB 21;  Length 128;
  Best Local Similarity   89.0%;  Pred. No. 2.1e-39;
  Matches   73;  Conservative    4;  Mismatches    5;  Indels    0;  Gaps    0;

Qy          1 MWLYRNPYVEAEYFPTKPMFVIAFLSPLSLIFLAKFLKKADTRDSRQACLAASLALALNG 60
              :|||||||||||||| ||||||||:|||||  ||||:|||  ||:|||||||||||||
Db         47 LWLYRNPYVEAEYFPTGPMFVIAFLTPLSLIFFAKFLRKADATDSKQACLAASLALALNG 106

Qy         61 VFTNTIKLIVGRPRPDFFYRCF 82
              |||| ||||||||||||||||
Db        107 VFTNIIKLIVGRPRPDFFYRCF 128


RESULT 11
AAB56000
ID   AAB56000 standard; Protein; 128 AA.
XX
AC   AAB56000;
XX
DT   08-MAR-2001  (first entry)
XX
DE   Skin cell protein, SEQ ID NO: 316.
XX
KW   Rat; skin cell; cytostatic; antiinflammatory; anti-HIV;
KW   nootropic; neuroprotective; vulnerary; immunomodulatory; vaccine;
KW   keratinocyte growth stimulation; cancer; angiogenesis inhibition;
KW   inflammation; neurological disease.
XX
OS   Rattus sp.
XX
PN   WO200069884-A2.
XX
PD   23-NOV-2000.
XX
PF   15-MAY-2000; 2000WO-NZ00075.
XX
PR   14-MAY-1999;   99US-0312283.
XX
PA   (GENE-) GENESIS RES & DEV CORP LTD.
XX
PI   Watson JD,  Strachan L,  Onrust R,  Sleeman M,  Kumble KD,  Murison JG;
XX
DR   WPI; 2001-007495/01.
DR   N-PSDB; AAC99699.
XX
PT   New isolated polynucleotide used in the identification of genetic
PT   disorders and encoding polypeptides used for treating inflammatory
PT   disease, cancer and neurological diseases -
XX
PS   Claim 4; Page 252-253; 352pp; English.
XX
CC   The present sequence is a polypeptide which is expressed in
```

```
CC   mammalian skin cells. The polypeptide is useful for stimulating
CC   keratinocyte growth and motility, inhibiting the growth of cancer cells,
CC   modulating angiogenesis, inhibiting angiogenesis and vascularisation of
CC   tumours, modulating skin inflammation, stimulating the growth of
CC   epithelial cells, inhibiting the binding of human immunodeficiency virus
CC   (HIV)-1 to leukocytes, and treating inflammmatory disease, cancer and
CC   neurological diseases. The polynucleotide can be used as a marker, in
CC   the identification of genetic disorders, and for the design of
CC   oligonucleotides for examining expression patterns.
XX
SQ   Sequence   128 AA;

  Query Match             40.2%;  Score 389;  DB 22;  Length 128;
  Best Local Similarity   89.0%;  Pred. No. 2.1e-39;
  Matches   73;  Conservative    4;  Mismatches    5;  Indels    0;  Gaps    0;

Qy          1 MWLYRNPYVEAEYFPTKPMFVIAFLSPLSLIFLAKFLKKADTRDSRQACLAASLALALNG 60
              :|||||||||||||||| ||||||||:|||||| ||||:||| ||:|||||||||||||
Db         47 LWLYRNPYVEAEYFPTGPMFVIAFLTPLSLIFFAKFLRKADATDSKQACLAASLALALNG 106

Qy         61 VFTNTIKLIVGRPRPDFFYRCF 82
              |||| ||||||||||||||||||
Db        107 VFTNIIKLIVGRPRPDFFYRCF 128


RESULT 12
ABB72200
ID   ABB72200 standard; Protein; 128 AA.
XX
AC   ABB72200;
XX
DT   04-APR-2002  (first entry)
XX
DE   Rat protein isolated from skin cells SEQ ID NO: 316.
XX
KW   Human; rat; mouse; skin cell; skin wound; cancer; growth defect;
KW   developmental defect; inflammatory disease; dermatological; vulnerary;
KW   immunomodulator; anti-inflammatory; cytostatic; neuroprotective.
XX
OS   Rattus sp.
XX
PN   WO200190357-A1.
XX
PD   29-NOV-2001.
XX
PF   24-MAY-2001; 2001WO-NZ00099.
XX
PR   24-MAY-2000; 2000US-206650P.
PR   25-JUL-2000; 2000US-221232P.
XX
PA   (GENE-) GENESIS RES & DEV CORP LTD.
XX
PI   Watson JD,  Strachan L,  Sleeman M,  Onrust R,  Murison JG,  Kumble KD;
XX
DR   WPI; 2002-122020/16.
XX
PT   New polynucleotides and polypeptides encoded by the polynucleotides
PT   isolated from skin cells, useful for treating skin wounds, cancers,
PT   growth and developmental defects, inflammatory diseases, or for
PT   modulating immune responses  -
XX
PS   Example 2; Page 208-209; 466pp; English.
XX
CC   The present invention provides the protein and coding sequences of cDNAs
CC   isolated from human, murine and rat skin cell libraries. The sequences
CC   can be used in the development of therapeutic agents useful in the
CC   treatment of skin diseases, including skin wounds, cancer, growth
CC   defects, developmental defects and inflammatory diseases. The proteins
CC   have important roles in the induction of hair growth, cell proliferation
CC   and cell-cell interaction, in maintaining tissue integrity, in wound
CC   healing and in modulating immune responses. The present sequence is a
CC   polypeptide of the invention.
XX
SQ   Sequence   128 AA;

  Query Match             40.2%;  Score 389;  DB 23;  Length 128;
  Best Local Similarity   89.0%;  Pred. No. 2.1e-39;
  Matches   73;  Conservative    4;  Mismatches    5;  Indels    0;  Gaps    0;

Qy          1 MWLYRNPYVEAEYFPTKPMFVIAFLSPLSLIFLAKFLKKADTRDSRQACLAASLALALNG 60
              :|||||||||||||||| ||||||||:|||||| ||||:||| ||:|||||||||||||
Db         47 LWLYRNPYVEAEYFPTGPMFVIAFLTPLSLIFFAKFLRKADATDSKQACLAASLALALNG 106

Qy         61 VFTNTIKLIVGRPRPDFFYRCF 82
              |||| ||||||||||||||||||
Db        107 VFTNIIKLIVGRPRPDFFYRCF 128


RESULT 13
AAG05044
ID   AAG05044 standard; Protein; 292 AA.
XX
AC   AAG05044;
XX
DT   17-OCT-2000  (first entry)
XX
DE   Arabidopsis thaliana protein fragment SEQ ID NO: 1300.
XX
KW   Protein identification; signal transduction pathway; metabolic pathway;
KW   hybridisation assay; genetic mapping; gene expression control; promoter;
KW   termination sequence.
XX
OS   Arabidopsis thaliana.
XX
PN   EP1033405-A2.
XX
PD   06-SEP-2000.
XX
PF   25-FEB-2000; 2000EP-0301439.
XX
PR   25-FEB-1999;   99US-0121825.
PR   05-MAR-1999;   99US-0123180.
PR   09-MAR-1999;   99US-0123548.
PR   23-MAR-1999;   99US-0125788.
PR   25-MAR-1999;   99US-0126264.
PR   29-MAR-1999;   99US-0126785.
PR   01-APR-1999;   99US-0127462.
PR   06-APR-1999;   99US-0128234.
PR   08-APR-1999;   99US-0128714.
PR   16-APR-1999;   99US-0129845.
PR   19-APR-1999;   99US-0130077.
PR   21-APR-1999;   99US-0130449.
PR   23-APR-1999;   99US-0130510.
PR   23-APR-1999;   99US-0130891.
PR   28-APR-1999;   99US-0131449.
PR   30-APR-1999;   99US-0132048.
PR   30-APR-1999;   99US-0132407.
PR   04-MAY-1999;   99US-0132484.
PR   05-MAY-1999;   99US-0132485.
PR   06-MAY-1999;   99US-0132486.
PR   06-MAY-1999;   99US-0132487.
PR   07-MAY-1999;   99US-0132863.
PR   11-MAY-1999;   99US-0134256.
PR   14-MAY-1999;   99US-0134218.
PR   14-MAY-1999;   99US-0134219.
PR   14-MAY-1999;   99US-0134221.
PR   14-MAY-1999;   99US-0134370.
PR   18-MAY-1999;   99US-0134768.
PR   19-MAY-1999;   99US-0134941.
PR   20-MAY-1999;   99US-0135124.
PR   21-MAY-1999;   99US-0135353.
PR   24-MAY-1999;   99US-0135629.
PR   25-MAY-1999;   99US-0136021.
PR   27-MAY-1999;   99US-0136392.
PR   28-MAY-1999;   99US-0136782.
PR   01-JUN-1999;   99US-0137222.
```

PR 03-JUN-1999; 99US-0137528.
PR 04-JUN-1999; 99US-0137502.
PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.
PR 10-JUN-1999; 99US-0138847.
PR 14-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 16-JUN-1999; 99US-0139453.
PR 17-JUN-1999; 99US-0139492.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139763.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161992.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.

Query Match 39.5%; Score 382; DB 21; Length 292;

```
  Best Local Similarity   47.0%;  Pred. No. 4.6e-38;
  Matches   77;  Conservative   28;  Mismatches   49;  Indels   10;  Gaps    4;

Qy         16 .TKPMF---VIAFLSPLSLIFLAKFLKKADTRDSRQACLAASLALALNGVFTNTIKLIVGR 72
                 | |::   | | | || :||:  : :: |  |   | |    :: :  | |: ||  |||
Db         43 TVPIWSVPVYAMLLPL-VIFIFIYFRRRDVYDLHHAVLGLLYSVLVTAVLTDAIKNAVGR 101

Qy         73 PRPDFFYRCFPDGLA----HSDLMCTGDKDVVNEGRKSFPSGHSSFAFAGLAFASFYLAG 128
              ||||||:|||||| |      |::| ||| |: || |||||||:|::|:|| | | ||:|
Db        102 PRPDFFWRCFPDGKALYDSLGDVICHGDKSVIREGHKSFPSGHTSWSFSGLGFLSLYLSG 161

Qy        129 KLHCFTPQGRGKSWRFCAFLSPLLFAAVIALSRTCDYKHHWQDL 172
              |:  |   |:|   : |  : ||||||:: :||  || |||||:
Db        162 KIQAF--DGKGHVAKLCIVILPLLFAALVGISRVDDYWHHWQDV 203


RESULT 14
AAG13844
ID   AAG13844 standard; Protein; 292 AA.
XX
AC   AAG13844;
XX
DT   17-OCT-2000  (first entry)
XX
DE   Arabidopsis thaliana protein fragment SEQ ID NO: 13485.
XX
KW   Protein identification; signal transduction pathway; metabolic pathway;
KW   hybridisation assay; genetic mapping; gene expression control; promoter;
KW   termination sequence.
XX
OS   Arabidopsis thaliana.
XX
PN   EP1033405-A2.
XX
PD   06-SEP-2000.
XX
PF   25-FEB-2000; 2000EP-0301439.
XX
PR   25-FEB-1999;   99US-0121825.
PR   05-MAR-1999;   99US-0123180.
PR   09-MAR-1999;   99US-0123548.
PR   23-MAR-1999;   99US-0125788.
PR   25-MAR-1999;   99US-0126264.
PR   29-MAR-1999;   99US-0126785.
PR   01-APR-1999;   99US-0127462.
PR   06-APR-1999;   99US-0128234.
PR   08-APR-1999;   99US-0128714.
PR   16-APR-1999;   99US-0129845.
PR   19-APR-1999;   99US-0130077.
PR   21-APR-1999;   99US-0130449.
PR   23-APR-1999;   99US-0130510.
PR   23-APR-1999;   99US-0130891.
PR   28-APR-1999;   99US-0131449.
PR   30-APR-1999;   99US-0132048.
PR   30-APR-1999;   99US-0132407.
PR   04-MAY-1999;   99US-0132484.
PR   05-MAY-1999;   99US-0132485.
PR   06-MAY-1999;   99US-0132486.
PR   06-MAY-1999;   99US-0132487.
PR   07-MAY-1999;   99US-0132863.
PR   11-MAY-1999;   99US-0134256.
PR   14-MAY-1999;   99US-0134218.
PR   14-MAY-1999;   99US-0134219.
PR   14-MAY-1999;   99US-0134221.
PR   14-MAY-1999;   99US-0134370.
PR   18-MAY-1999;   99US-0134768.
PR   19-MAY-1999;   99US-0134941.
PR   20-MAY-1999;   99US-0135124.
PR   21-MAY-1999;   99US-0135353.
PR   24-MAY-1999;   99US-0135629.
PR   25-MAY-1999;   99US-0136021.
PR   27-MAY-1999;   99US-0136392.
PR   28-MAY-1999;   99US-0136782.
PR   01-JUN-1999;   99US-0137222.
PR   03-JUN-1999;   99US-0137528.
PR   04-JUN-1999;   99US-0137502.
PR   07-JUN-1999;   99US-0137724.
PR   08-JUN-1999;   99US-0138094.
PR   10-JUN-1999;   99US-0138540.
PR   10-JUN-1999;   99US-0138847.
PR   14-JUN-1999;   99US-0139119.
PR   16-JUN-1999;   99US-0139452.
PR   16-JUN-1999;   99US-0139453.
PR   17-JUN-1999;   99US-0139492.
PR   18-JUN-1999;   99US-0139454.
PR   18-JUN-1999;   99US-0139455.
PR   18-JUN-1999;   99US-0139456.
PR   18-JUN-1999;   99US-0139457.
PR   18-JUN-1999;   99US-0139458.
PR   18-JUN-1999;   99US-0139459.
PR   18-JUN-1999;   99US-0139460.
PR   18-JUN-1999;   99US-0139461.
PR   18-JUN-1999;   99US-0139462.
PR   18-JUN-1999;   99US-0139463.
PR   18-JUN-1999;   99US-0139750.
PR   18-JUN-1999;   99US-0139763.
PR   21-JUN-1999;   99US-0139817.
PR   22-JUN-1999;   99US-0139899.
PR   23-JUN-1999;   99US-0140353.
PR   23-JUN-1999;   99US-0140354.
PR   24-JUN-1999;   99US-0140695.
PR   28-JUN-1999;   99US-0140823.
PR   29-JUN-1999;   99US-0140991.
PR   30-JUN-1999;   99US-0141287.
PR   01-JUL-1999;   99US-0141842.
PR   01-JUL-1999;   99US-0142154.
PR   02-JUL-1999;   99US-0142055.
PR   06-JUL-1999;   99US-0142390.
PR   08-JUL-1999;   99US-0142803.
PR   09-JUL-1999;   99US-0142920.
PR   12-JUL-1999;   99US-0142977.
PR   13-JUL-1999;   99US-0143542.
PR   14-JUL-1999;   99US-0143624.
PR   15-JUL-1999;   99US-0144005.
PR   16-JUL-1999;   99US-0144085.
PR   16-JUL-1999;   99US-0144086.
PR   19-JUL-1999;   99US-0144325.
PR   19-JUL-1999;   99US-0144331.
PR   19-JUL-1999;   99US-0144332.
PR   19-JUL-1999;   99US-0144333.
PR   19-JUL-1999;   99US-0144334.
PR   19-JUL-1999;   99US-0144335.
PR   20-JUL-1999;   99US-0144352.
PR   20-JUL-1999;   99US-0144632.
PR   20-JUL-1999;   99US-0144884.
PR   21-JUL-1999;   99US-0144814.
PR   21-JUL-1999;   99US-0145086.
PR   21-JUL-1999;   99US-0145088.
PR   22-JUL-1999;   99US-0145085.
PR   22-JUL-1999;   99US-0145087.
PR   22-JUL-1999;   99US-0145089.
PR   22-JUL-1999;   99US-0145192.
PR   23-JUL-1999;   99US-0145145.
PR   23-JUL-1999;   99US-0145218.
PR   23-JUL-1999;   99US-0145224.
PR   26-JUL-1999;   99US-0145276.
PR   27-JUL-1999;   99US-0145913.
PR   27-JUL-1999;   99US-0145918.
PR   27-JUL-1999;   99US-0145919.
PR   28-JUL-1999;   99US-0145951.
PR   02-AUG-1999;   99US-0146386.
PR   02-AUG-1999;   99US-0146388.
PR   02-AUG-1999;   99US-0146389.
PR   03-AUG-1999;   99US-0147038.
PR   04-AUG-1999;   99US-0147204.
PR   04-AUG-1999;   99US-0147302.
```

PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161992.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.

Query Match 39.5%; Score 382; DB 21; Length 292;
Best Local Similarity 47.0%; Pred. No. 4.6e-38;
Matches 77; Conservative 28; Mismatches 49; Indels 10; Gaps 4;

```
Qy         16 TKPMF---VIAFLSPLSLIFLAKFLKKADTRDSRQACLAASLALALNGVFTNTIKLIVGR 72
              | |::   | | | || :||:  : :: |  |   | |    :: :  | |: ||  |||
Db         43 TVPIWSVPVYAMLLPL-VIFIFIYFRRRDVYDLHHAVLGLLLYSVLVTAVLTDAIKNAVGR 101

Qy         73 PRPDFFYRCFPDGLA----HSDLMCTGDKDVVNEGRKSFPSGHSSFAFAGLAFASFYLAG 128
              ||||||:|||||| |      |::| ||| |: || |||||||:|::|:|| | | ||:|
Db        102 PRPDFFWRCFPDGKALYDSLGDVICHGDKSVIREGHKSFPSGHTSWSFSGLGFLSLYLSG 161

Qy        129 KLHCFTPQGRGKSWRFCAFLSPLLFAAVIALSRTCDYKHHWQDL 172
              |:  |   |:|   : |  : ||||||:: :||  || |||||:
Db        162 KIQAF--DGKGHVAKLCIVILPLLFAALVGISRVDDYWHHWQDV 203
```

RESULT 15
AAG14938
ID AAG14938 standard; Protein; 292 AA.
XX
AC AAG14938;
XX
DT 17-OCT-2000 (first entry)
XX
DE Arabidopsis thaliana protein fragment SEQ ID NO: 14988.
XX
KW Protein identification; signal transduction pathway; metabolic pathway;
KW hybridisation assay; genetic mapping; gene expression control; promoter;
KW termination sequence.
XX
OS Arabidopsis thaliana.
XX
PN EP1033405-A2.
XX
PD 06-SEP-2000.
XX
PF 25-FEB-2000; 2000EP-0301439.
XX
PR 25-FEB-1999; 99US-0121825.
PR 05-MAR-1999; 99US-0123180.
PR 09-MAR-1999; 99US-0123548.
PR 23-MAR-1999; 99US-0125788.
PR 25-MAR-1999; 99US-0126264.
PR 29-MAR-1999; 99US-0126785.
PR 01-APR-1999; 99US-0127462.
PR 06-APR-1999; 99US-0128234.
PR 08-APR-1999; 99US-0128714.
PR 16-APR-1999; 99US-0129845.
PR 19-APR-1999; 99US-0130077.
PR 21-APR-1999; 99US-0130449.
PR 23-APR-1999; 99US-0130510.
PR 23-APR-1999; 99US-0130891.
PR 28-APR-1999; 99US-0131449.
PR 30-APR-1999; 99US-0132048.
PR 30-APR-1999; 99US-0132407.
PR 04-MAY-1999; 99US-0132484.
PR 05-MAY-1999; 99US-0132485.
PR 06-MAY-1999; 99US-0132486.
PR 06-MAY-1999; 99US-0132487.
PR 07-MAY-1999; 99US-0132863.
PR 11-MAY-1999; 99US-0134256.
PR 14-MAY-1999; 99US-0134218.
PR 14-MAY-1999; 99US-0134219.
PR 14-MAY-1999; 99US-0134221.
PR 14-MAY-1999; 99US-0134370.
PR 18-MAY-1999; 99US-0134768.
PR 19-MAY-1999; 99US-0134941.
PR 20-MAY-1999; 99US-0135124.
PR 21-MAY-1999; 99US-0135353.
PR 24-MAY-1999; 99US-0135629.
PR 25-MAY-1999; 99US-0136021.
PR 27-MAY-1999; 99US-0136392.

PR 28-MAY-1999; 99US-0136782.
PR 01-JUN-1999; 99US-0137222.
PR 03-JUN-1999; 99US-0137528.
PR 04-JUN-1999; 99US-0137502.
PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.
PR 10-JUN-1999; 99US-0138847.
PR 14-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 16-JUN-1999; 99US-0139453.
PR 17-JUN-1999; 99US-0139492.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139763.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161992.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.

Query Match 39.5%; Score 382; DB 21; Length 292;
Best Local Similarity 47.0%; Pred. No. 4.6e-38;
Matches 77; Conservative 28; Mismatches 49; Indels 10; Gaps 4;

```
Qy          16 TKPMF---VIAFLSPLSLIFLAKFLKKADTRDSRQACLAASLALALNGVFTNTIKLIVGR 72
               | |::   | | | || :||:  : :: |  |   | |    :: :  | |: ||  |||
Db          43 TVPIWSVPVYAMLLPL-VIFIFIYFRRRDVYDLHHAVLGLLYSVLVTAVLTDAIKNAVGR 101

Qy          73 PRPDFFYRCFPDGLA----HSDLMCTGDKDVVNEGRKSFPSGHSSFAFAGLAFASFYLAG 128
               ||||||:|||||| |      |::| ||| |: || |||||||:|::|:|| | | ||:|
Db         102 PRPDFFWRCFPDGKALYDSLGDVICHGDKSVIREGHKSFPSGHTSWSFSGLGFLSLYLSG 161

Qy         129 KLHCFTPQGRGKSWRFCAFLSPLLFAAVIALSRTCDYKHHWQDL 172
               |:  |   |:|   : |  : ||||||:: :||  ||. |||||:
Db         162 KIQAF--DGKGHVAKLCIVILPLLFAALVGISRVDDYWHHWQDV 203
```

Search completed: August 14, 2003, 13:05:41
Job time : 87 secs

GenCore version 5.1.6

OM protein - protein search, using sw model

Run on: August 14, 2003, 13:03:37 ; Search time 30 Seconds
(without alignments)
253.865 Million cell updates/sec

Title: US-09-876-997-399
Perfect score: 968
Sequence: 1 MWLYRNPYVEAEYFPTKPMF.....RTCDYKHHWQDLLKCTNTAK 180

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 328717 seqs, 42310858 residues

Total number of hits satisfying chosen parameters: 328717

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued_Patents_AA:*
1: /cgn2_6/ptodata/1/iaa/5A_COMB.pep:*
2: /cgn2_6/ptodata/1/iaa/5B_COMB.pep:*
3: /cgn2_6/ptodata/1/iaa/6A_COMB.pep:*
4: /cgn2_6/ptodata/1/iaa/6B_COMB.pep:*
5: /cgn2_6/ptodata/1/iaa/PCTUS_COMB.pep:*
6: /cgn2_6/ptodata/1/iaa/backfiles1.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	389	40.2	128	3	US-09-188-930-316	Sequence 316, App
2	389	40.2	128	4	US-09-312-283C-316	Sequence 316, App
3	382	39.5	314	4	US-09-360-376-14	Sequence 14, Appl
4	377	38.9	322	4	US-09-360-376-16	Sequence 16, Appl
5	372	38.4	290	4	US-09-360-376-12	Sequence 12, Appl
6	357	36.9	343	4	US-09-360-376-17	Sequence 17, Appl
7	337	34.8	348	4	US-09-360-376-13	Sequence 13, Appl
8	335	34.6	310	4	US-09-360-376-15	Sequence 15, Appl
9	247	25.5	289	4	US-09-360-376-52	Sequence 52, Appl
10	203.5	21.0	123	3	US-09-188-930-170	Sequence 170, App
11	203.5	21.0	123	4	US-09-312-283C-170	Sequence 170, App
12	184.5	19.1	282	4	US-09-360-376-54	Sequence 54, Appl
13	183.5	19.0	283	3	US-08-992-035A-3	Sequence 3, Appli
14	183.5	19.0	283	4	US-09-360-376-53	Sequence 53, Appl
15	174.5	18.0	285	3	US-08-992-035A-1	Sequence 1, Appli
16	127	13.1	233	4	US-09-360-376-55	Sequence 55, Appl
17	114	11.8	323	4	US-09-122-315C-18	Sequence 18, Appl
18	80	8.3	199	4	US-09-107-532A-7261	Sequence 7261, Ap
19	76	7.9	244	4	US-09-252-991A-21937	Sequence 21937, A
20	74	7.6	214	4	US-09-328-352-4246	Sequence 4246, Ap
21	72.5	7.5	401	4	US-09-252-991A-27198	Sequence 27198, A
22	72.5	7.5	1765	4	US-09-354-147C-5	Sequence 5, Appli
23	72	7.4	444	4	US-09-328-352-5249	Sequence 5249, Ap
24	71.5	7.4	1765	4	US-09-354-147C-2	Sequence 2, Appli
25	71.5	7.4	1765	4	US-09-354-147C-3	Sequence 3, Appli
26	70.5	7.3	254	4	US-09-107-532A-4025	Sequence 4025, Ap
27	70	7.2	202	4	US-09-252-991A-26505	Sequence 26505, A
28	70	7.2	235	3	US-09-248-335-64	Sequence 64, Appl
29	69	7.1	1684	3	US-08-665-259-25	Sequence 25, Appl
30	69	7.1	1684	3	US-08-762-500-25	Sequence 25, Appl
31	69	7.1	1704	3	US-08-762-500-75	Sequence 75, Appl
32	69	7.1	2233	2	US-08-569-853-1	Sequence 1, Appli
33	69	7.1	2233	2	US-08-569-853-2	Sequence 2, Appli
34	69	7.1	2233	3	US-08-987-439-1	Sequence 1, Appli
35	67.5	7.0	429	4	US-09-351-150A-5	Sequence 5, Appli
36	67.5	7.0	432	4	US-09-252-991A-22739	Sequence 22739, A
37	67.5	7.0	444	4	US-08-482-746-15	Sequence 15, Appl
38	67.5	7.0	444	4	US-09-580-734-15	Sequence 15, Appl
39	67.5	7.0	444	4	US-08-374-009-15	Sequence 15, Appl
40	67	6.9	308	4	US-09-328-352-4190	Sequence 4190, Ap
41	67	6.9	940	4	US-09-198-452A-500	Sequence 500, App
42	66.5	6.9	233	4	US-09-134-001C-4013	Sequence 4013, Ap
43	65.5	6.8	315	4	US-09-674-529B-15	Sequence 15, Appl
44	65.5	6.8	1968	1	US-07-745-206A-7	Sequence 7, Appli
45	65.5	6.8	1968	1	US-08-455-543A-45	Sequence 45, Appl

ALIGNMENTS

RESULT 1
US-09-188-930-316
; Sequence 316, Application US/09188930A
; Patent No. 6150502
; GENERAL INFORMATION:
; APPLICANT: Watson, James D.
; APPLICANT: Strachan, Lorna
; APPLICANT: Sleeman, Matthew
; APPLICANT: Onrust, Rene
; APPLICANT: Murison, James Greg
; TITLE OF INVENTION: Compositions Isolated From Skin Cells
; TITLE OF INVENTION: and Methods For Their Use
; FILE REFERENCE: 11000.1011c1
; CURRENT APPLICATION NUMBER: US/09/188,930A
; CURRENT FILING DATE: 1998-11-09
; NUMBER OF SEQ ID NOS: 348
; SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 316
; LENGTH: 128
; TYPE: PRT
; ORGANISM: Rat
US-09-188-930-316

Query Match 40.2%; Score 389; DB 3; Length 128;
Best Local Similarity 89.0%; Pred. No. 7e-43;
Matches 73; Conservative 4; Mismatches 5; Indels 0; Gaps 0;

```
Qy          1 MWLYRNPYVEAEYFPTKPMFVIAFLSPLSLIFLAKFLKKADTRDSRQACLAASLALALNG 60
              :|||||||||||||| ||||||||:|||||| ||||:|||  ||:|||||||||||||
Db         47 LWLYRNPYVEAEYFPTGPMFVIAFLTPLSLIFFAKFLRKADATDSKQACLAASLALALNG 106

Qy         61 VFTNTIKLIVGRPRPDFFYRCF 82
              |||| ||||||||||||||||||
Db        107 VFTNIIKLIVGRPRPDFFYRCF 128
```

RESULT 2
US-09-312-283C-316
; Sequence 316, Application US/09312283C
; Patent No. 6573095
; GENERAL INFORMATION:
; APPLICANT: Watson, James D.
; APPLICANT: Strachan, Lorna
; APPLICANT: Sleeman, Matthew
; APPLICANT: Onrust, Rene
; APPLICANT: Murison, James G.
; APPLICANT: Kumble, Krishanand D.
; TITLE OF INVENTION: Compositions Isolated from Skin Cells
; TITLE OF INVENTION: and Methods for Their Use
; FILE REFERENCE: 11000.1011c2

```
;  CURRENT APPLICATION NUMBER: US/09/312,283C
;  CURRENT FILING DATE:  1999-05-14
;  NUMBER OF SEQ ID NOS: 425
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 316
;   LENGTH: 128
;   TYPE: PRT
;   ORGANISM: Mouse
US-09-312-283C-316

  Query Match             40.2%;  Score 389;  DB 4;  Length 128;
  Best Local Similarity   89.0%;  Pred. No. 7e-43;
  Matches   73;  Conservative    4;  Mismatches    5;  Indels    0;  Gaps    0;

Qy          1 MWLYRNPYVEAEYFPTKPMFVIAFLSPLSLIFLAKFLKKADTRDSRQACLAASLALALNG 60
              :|||||||||||||||| ||||||||:|||||| ||||:|||  ||:|||||||||||||
Db         47 LWLYRNPYVEAEYFPTGPMFVIAFLTPLSLIFFAKFLRKADATDSKQACLAASLALALNG 106

Qy         61 VFTNTIKLIVGRPRPDFFYRCF 82
              |||| ||||||||||||||||
Db        107 VFTNIIKLIVGRPRPDFFYRCF 128


RESULT 3
US-09-360-376-14
; Sequence 14, Application US/09360376
; Patent No. 6495739
; GENERAL INFORMATION:
;  APPLICANT:  Lassner, Michael
;  APPLICANT:  Ruezinsky, Diane
;  TITLE OF INVENTION:   PLANT PHOSPHATIDIC ACID PHOSPHATASES
;  FILE REFERENCE:   17026/01/US
;  CURRENT APPLICATION NUMBER: US/09/360,376
;  CURRENT FILING DATE:  1999-07-23
;  PRIOR APPLICATION NUMBER:   US 09/122,315
;  PRIOR FILING DATE:   1998-07-24
;  NUMBER OF SEQ ID NOS:   55
;  SOFTWARE:   PatentIn version 3.0
; SEQ ID NO 14
;   LENGTH: 314
;   TYPE: PRT
;   ORGANISM: Arabidopsis sp.
US-09-360-376-14

  Query Match             39.5%;  Score 382;  DB 4;  Length 314;
  Best Local Similarity   47.0%;  Pred. No. 2e-41;
  Matches   77;  Conservative   28;  Mismatches   49;  Indels   10;  Gaps    4;

Qy         16 TKPMF---VIAFLSPLSLIFLAKFLKKADTRDSRQACLAASLALALNGVFTNTIKLIVGR 72
              | |::   | | | || :||:  : :: |  |   | |    :: :  | |: ||  |||
Db         65 TVPIWSVPVYAMLLPL-VIFIFIYFRRRDVYDLHHAVLGLLYSVLVTAVLTDAIKNAVGR 123

Qy         73 PRPDFFYRCFPDGLA----HSDLMCTGDKDVVNEGRKSFPSGHSSFAFAGLAFASFYLAG 128
              ||||||:|||||| |       |::| ||| |: || ||||||||:|::|:|| | | ||:|
Db        124 PRPDFFWRCFPDGKALYDSLGDVICHGDKSVIREGHKSFPSGHTSWSFSGLGFLSLYLSG 183

Qy        129 KLHCFTPQGRGKSWRFCAFLSPLLFAAVIALSRTCDYKHHWQDL 172
              |:  |   |:|   : |  : ||||||:: :||  || |||||:
Db        184 KIQAF--DGKGHVAKLCIVILPLLFAALVGISRVDDYWHHWQDV 225


RESULT 4
US-09-360-376-16
; Sequence 16, Application US/09360376
; Patent No. 6495739
; GENERAL INFORMATION:
;  APPLICANT:  Lassner, Michael
;  APPLICANT:  Ruezinsky, Diane
;  TITLE OF INVENTION:   PLANT PHOSPHATIDIC ACID PHOSPHATASES
;  FILE REFERENCE:   17026/01/US
;  CURRENT APPLICATION NUMBER: US/09/360,376
;  CURRENT FILING DATE:  1999-07-23
;  PRIOR APPLICATION NUMBER:   US 09/122,315
;  PRIOR FILING DATE:   1998-07-24
;  NUMBER OF SEQ ID NOS:   55
;  SOFTWARE:   PatentIn version 3.0
; SEQ ID NO 16
;   LENGTH: 322
;   TYPE: PRT
;   ORGANISM: Glycine sp.
US-09-360-376-16

  Query Match             38.9%;  Score 377;  DB 4;  Length 322;
  Best Local Similarity   45.0%;  Pred. No. 9.6e-41;
  Matches   76;  Conservative   31;  Mismatches   54;  Indels    8;  Gaps    3;

Qy          9 VEAEYFPTKPMFVIAFLSPLSLIFLAKFLKKADTRDSRQACLAASLALALNGVFTNTIKL 68
              ::|    |   : :|| | ||::  :  |::| |  |   | :    :: :  | |: ||
Db         61 LKANTIPFWAVPIIAILLPLAVFLVYYFIRK-DVYDLHHAIMGLLFSVLITAVMTDAIKD 119

Qy         69 IVGRPRPDFFYRCFPDGL-----AHSDLMCTGDKDVVNEGRKSFPSGHSSFAFAGLAFAS 123
              |||||||||:||||||        |:::||||| |: || |||||||:|::|||| : :
Db        120 AVGRPRPDFFWRCFPDGKGVFDPVTSNVLCTGDKGVIKEGHKSFPSGHTSWSFAGLVYLA 179

Qy        124 FYLAGKLHCFTPQGRGKSWRFCAFLSPLLFAAVIALSRTCDYKHHWQDL 172
              :||:|||  |    ||   : |    |:| ||:||:||  || |||||:
Db        180 WYLSGKLRAF--DRRGHVAKLCLVFLPILVAAMIAVSRVDDYWHHWQDV 226


RESULT 5
US-09-360-376-12
; Sequence 12, Application US/09360376
; Patent No. 6495739
; GENERAL INFORMATION:
;  APPLICANT:  Lassner, Michael
;  APPLICANT:  Ruezinsky, Diane
;  TITLE OF INVENTION:   PLANT PHOSPHATIDIC ACID PHOSPHATASES
;  FILE REFERENCE:   17026/01/US
;  CURRENT APPLICATION NUMBER: US/09/360,376
;  CURRENT FILING DATE:  1999-07-23
;  PRIOR APPLICATION NUMBER:   US 09/122,315
;  PRIOR FILING DATE:   1998-07-24
;  NUMBER OF SEQ ID NOS:   55
;  SOFTWARE:   PatentIn version 3.0
; SEQ ID NO 12
;   LENGTH: 290
;   TYPE: PRT
;   ORGANISM: Arabidopsis sp.
US-09-360-376-12

  Query Match             38.4%;  Score 372;  DB 4;  Length 290;
  Best Local Similarity   47.8%;  Pred. No. 3.7e-40;
  Matches   75;  Conservative   28;  Mismatches   46;  Indels    8;  Gaps    4;

Qy         21 VIAFLSPLSLIFLAKFLKKADTRDSRQACLAASLALALNGVFTNTIKLIVGRPRPDFFYR 80
              :|| : | ::| :  |::  |  |    | |     :: : || |: ||  |||||||||:|
Db         73 LIAVVLPFAVICVYYFIRN-DVYDLHHAILGLLFSVLITGVITDAIKDAVGRPRPDFFWR 131

Qy         81 CFPDGLA--HS---DLMCTGDKDVVNEGRKSFPSGHSSFAFAGLAFASFYLAGKLHCFTP 135
              |||||:   |:   :::||| |||| || |||||||:|::|||| | | ||:|||:  |
Db        132 CFPDGIGIFHNVTKNVLCTGAKDVVKEGHKSFPSGHTSWSFAGLGFLSLYLSGKIRVF-- 189

Qy        136 QGRGKSWRFCAFLSPLLFAAVIALSRTCDYKHHWQDL 172
              ||   : |  : ||| ||:: :||  || |||||:
Db        190 DQRGHVAKLCIVILPLLVAALVGVSRVDDYWHHWQDV 226


RESULT 6
US-09-360-376-17
; Sequence 17, Application US/09360376
; Patent No. 6495739
; GENERAL INFORMATION:
;  APPLICANT:  Lassner, Michael
;  APPLICANT:  Ruezinsky, Diane
;  TITLE OF INVENTION:   PLANT PHOSPHATIDIC ACID PHOSPHATASES
;  FILE REFERENCE:   17026/01/US
```

```
;  CURRENT APPLICATION NUMBER: US/09/360,376
;  CURRENT FILING DATE:  1999-07-23
;  PRIOR APPLICATION NUMBER:   US 09/122,315
;  PRIOR FILING DATE:   1998-07-24
;  NUMBER OF SEQ ID NOS:   55
;  SOFTWARE:   PatentIn version 3.0
; SEQ ID NO 17
;   LENGTH: 343
;   TYPE: PRT
;   ORGANISM: Glycine sp.
US-09-360-376-17

  Query Match            36.9%;  Score 357;  DB 4;  Length 343;
  Best Local Similarity  44.5%;  Pred. No. 4.3e-38;
  Matches   73;  Conservative   27;  Mismatches   54;  Indels   10;  Gaps    4;

Qy         16 TKPMFVI---AFLSPLSLIFLAKFLKKADTRDSRQACLAASLALALNGVFTNTIKLIVGR 72
              | | : |   | | |: :|||  :::: |  |   | |    :: :  |||  ||  |||
Db         89 TVPAWAIPIYAILLPI-VIFLGVYIRRRDVYDLHHAVLGLLFSVLITAVFTEAIKNAVGR 147

Qy         73 PRPDFFYRCFPDGL----AHSDLMCTGDKDVVNEGRKSFPSGHSSFAFAGLAFASFYLAG 128
              ||||||:||||||        |::| ||: |: || |||||||:| :|:|| | | ||:|
Db        148 PRPDFFWRCFPDGKDVYDKWGDVICHGDQKVIKEGYKSFPSGHTSGSFSGLGFLSLYLSG 207

Qy        129 KLHCFTPQGRGKSWRFCAFLSPLLFAAVIALSRTCDYKHHWQDL 172
              |: |   :|   : |    ||| |::: :||  || |||||:
Db        208 KIKAF--DRKGHVAKLCIVFLPLLVASLVGISRVDDYWHHWQDV 249


RESULT 7
US-09-360-376-13
; Sequence 13, Application US/09360376
; Patent No. 6495739
; GENERAL INFORMATION:
;  APPLICANT:  Lassner, Michael
;  APPLICANT:  Ruezinsky, Diane
;  TITLE OF INVENTION:   PLANT PHOSPHATIDIC ACID PHOSPHATASES
;  FILE REFERENCE:   17026/01/US
;  CURRENT APPLICATION NUMBER: US/09/360,376
;  CURRENT FILING DATE:  1999-07-23
;  PRIOR APPLICATION NUMBER:   US 09/122,315
;  PRIOR FILING DATE:   1998-07-24
;  NUMBER OF SEQ ID NOS:   55
;  SOFTWARE:   PatentIn version 3.0
; SEQ ID NO 13
;   LENGTH: 348
;   TYPE: PRT
;   ORGANISM: Arabidopsis sp.
US-09-360-376-13

  Query Match            34.8%;  Score 337;  DB 4;  Length 348;
  Best Local Similarity  42.1%;  Pred. No. 1.8e-35;
  Matches   69;  Conservative   31;  Mismatches   54;  Indels   10;  Gaps    4;

Qy         16 TKPMF---VIAFLSPLSLIFLAKFLKKADTRDSRQACLAASLALALNGVFTNTIKLIVGR 72
              | |::   | | | |: ::|:  :||:    |   : |    |: : || |::||:  ||
Db        111 TVPIWSVPVYAVLLPI-IVFVCFYLKRTCVYDLHHSILGLLFAVLITGVITDSIKVATGR 169

Qy         73 PRPDFFYRCFPDGL----AHSDLMCTGDKDVVNEGRKSFPSGHSSFAFAGLAFASFYLAG 128
              |||:|::||||||     |   ::| |    | || |||||||:|::|||| | | ||:|
Db        170 PRPNFYWRCFPDGKELYDALGGVVCHGKAAEVKEGHKSFPSGHTSWSFAGLTFLSLYLSG 229

Qy        129 KLHCFTPQGRGKSWRFCAFLSPLLFAAVIALSRTCDYKHHWQDL 172
              |:  |  :|     : |  : ||| | :: :||  || |||||:
Db        230 KIKAFNNEGHVA--KLCLVIFPLLAACLVGISRVDDYWHHWQDV 271


RESULT 8
US-09-360-376-15
; Sequence 15, Application US/09360376
; Patent No. 6495739
; GENERAL INFORMATION:
;  APPLICANT:  Lassner, Michael
;  APPLICANT:  Ruezinsky, Diane
;  TITLE OF INVENTION:   PLANT PHOSPHATIDIC ACID PHOSPHATASES
;  FILE REFERENCE:   17026/01/US
;  CURRENT APPLICATION NUMBER: US/09/360,376
;  CURRENT FILING DATE:  1999-07-23
;  PRIOR APPLICATION NUMBER:   US 09/122,315
;  PRIOR FILING DATE:   1998-07-24
;  NUMBER OF SEQ ID NOS:   55
;  SOFTWARE:   PatentIn version 3.0
; SEQ ID NO 15
;   LENGTH: 310
;   TYPE: PRT
;   ORGANISM: Zea mays
US-09-360-376-15

  Query Match            34.6%;  Score 335;  DB 4;  Length 310;
  Best Local Similarity  45.2%;  Pred. No. 2.8e-35;
  Matches   71;  Conservative   25;  Mismatches   53;  Indels    8;  Gaps    3;

Qy         21 VIAFLSPLSLIFLAKFLKKADTRDSRQACLAASLALALNGVFTNTIKLIVGRPRPDFFYR 80
              :|  : | : ||:  : || :  |      |    :: :  | |: ||  |||||||||:|
Db         72 LIGIILPWA-IFVGIYFKKKNFYDLHHGILGILYSVLITAVITDAIKDGVGRPRPDFFWR 130

Qy         81 CFPDGLAHSD-----LMCTGDKDVVNEGRKSFPSGHSSFAFAGLAFASFYLAGKLHCFTP 135
              |||:|    |      ::| | | |: || ||||||||::|||| | ::||||||  |
Db        131 CFPNGNDVYDNITTGVICNGVKSVIKEGHKSFPSGHSSWSFAGLGFLAWYLAGKLTAF-- 188

Qy        136 QGRGKSWRFCAFLSPLLFAAVIALSRTCDYKHHWQDL 172
              :|   : |    ||| ||::|:||  || |||||:
Db        189 DRKGHIAKLCIVFLPLLTAALVAVSRVDDYWHHWQDV 225


RESULT 9
US-09-360-376-52
; Sequence 52, Application US/09360376
; Patent No. 6495739
; GENERAL INFORMATION:
;  APPLICANT:  Lassner, Michael
;  APPLICANT:  Ruezinsky, Diane
;  TITLE OF INVENTION:   PLANT PHOSPHATIDIC ACID PHOSPHATASES
;  FILE REFERENCE:   17026/01/US
;  CURRENT APPLICATION NUMBER: US/09/360,376
;  CURRENT FILING DATE:  1999-07-23
;  PRIOR APPLICATION NUMBER:   US 09/122,315
;  PRIOR FILING DATE:   1998-07-24
;  NUMBER OF SEQ ID NOS:   55
;  SOFTWARE:   PatentIn version 3.0
; SEQ ID NO 52
;   LENGTH: 289
;   TYPE: PRT
;   ORGANISM: Saccharomyces cerevisiae
US-09-360-376-52

  Query Match            25.5%;  Score 247;  DB 4;  Length 289;
  Best Local Similarity  37.8%;  Pred. No. 8.1e-24;
  Matches   68;  Conservative   27;  Mismatches   69;  Indels   16;  Gaps    6;

Qy          6 NPYVEAEYFPTKPMFVIAFLSP-LSLIFLAKFLKKADTRD----SRQACLAASLALALNG 60
              :||   |      :|| :|: | |::: :   |  || |         : |  |||
Db         54 HPYATTERVNNNMLFVYSFVVPSLTILIIGSIL--ADRRHLIFILYTSLLGLSLAWFSTS 111

Qy         61 VFTNTIKLIVGRPRPDFFYRCFP-DGL------AHSDLMCTGDKDVVNEGRKSFPSGHSS 113
              ||| ||  :|| ||||  || | :||         |:  | : : : :| :: ||||||
Db        112 FFTNFIKNWIGRLRPDFLDRCQPVEGLPLDTLFTAKDVCTTKNHERLLDGFRTTPSGHSS 171

Qy        114 FAFAGLAFASFYLAGKLHCFTPQGRGKSWRFCAFLSPLLFAAVIALSRTCDYKHHWQDLL 173
              :|||| :  |:| |:|   |        ||      ||| ||:|||||| ||:||: |::
Db        172 ESFAGLGYLYFWLCGQL--LTESPLMPLWRKMVAFLPLLGAALIALSRTQDYRHHFVDVI 229


RESULT 10
US-09-188-930-170
; Sequence 170, Application US/09188930A
```

```
; Patent No. 6150502
; GENERAL INFORMATION:
;  APPLICANT: Watson, James D.
;  APPLICANT: Strachan, Lorna
;  APPLICANT: Sleeman, Matthew
;  APPLICANT: Onrust, Rene
;  APPLICANT: Murison, James Greg
;  TITLE OF INVENTION: Compositions Isolated From Skin Cells
;  TITLE OF INVENTION: and Methods For Their Use
;  FILE REFERENCE: 11000.1011c1
;  CURRENT APPLICATION NUMBER: US/09/188,930A
;  CURRENT FILING DATE: 1998-11-09
;  NUMBER OF SEQ ID NOS: 348
;  SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 170
;   LENGTH: 123
;   TYPE: PRT
;   ORGANISM: Rat
;   FEATURE:
;   NAME/KEY: UNSURE
;   LOCATION: (27)...(27)
;   NAME/KEY: UNSURE
;   LOCATION: (104)...(104)
;   NAME/KEY: UNSURE
;   LOCATION: (118)...(118)
US-09-188-930-170

  Query Match             21.0%;  Score 203.5;  DB 3;  Length 123;
  Best Local Similarity   45.2%;  Pred. No. 1.2e-18;
  Matches   47;  Conservative    8;  Mismatches   20;  Indels   29;  Gaps    3;

Qy          1 MWLYRNPYVEAEYFPTKPMFVIAFLSPLSLIFLAKFLKKADTRDSR------QACLAASL 54
              :|||||||:|||||| ||||||||:|||||| ||||:|||     |       | | :|
Db         34 LWLYRNPYVKAEYFPTGPMFVIAFLTPLSLIFFAKFLRKADADRQRASLPRCQPCPSAKW 93

Qy         55 ALAL---------NGVFTNTIKLIVGRPRPDFFYRCFPDGLAHS 89
                |         :    :|              | | |:|||
Db         94 CLYQHHKTDSXQGHAQIAST-------------ECSPXGIAHS 123


RESULT 11
US-09-312-283C-170
; Sequence 170, Application US/09312283C
; Patent No. 6573095
; GENERAL INFORMATION:
;  APPLICANT: Watson, James D.
;  APPLICANT:  Strachan, Lorna
;  APPLICANT:  Sleeman, Matthew
;  APPLICANT:  Onrust, Rene
;  APPLICANT:  Murison, James G.
;  APPLICANT:  Kumble, Krishanand D.
;  TITLE OF INVENTION: Compositions Isolated from Skin Cells
;  TITLE OF INVENTION:  and Methods for Their Use
;  FILE REFERENCE: 11000.1011c2
;  CURRENT APPLICATION NUMBER: US/09/312,283C
;  CURRENT FILING DATE:  1999-05-14
;  NUMBER OF SEQ ID NOS: 425
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 170
;   LENGTH: 123
;   TYPE: PRT
;   ORGANISM: Mouse
;   FEATURE:
;   NAME/KEY: VARIANT
;   LOCATION: (1)...(123)
;   OTHER INFORMATION: Xaa = Any Amino Acid
US-09-312-283C-170

  Query Match             21.0%;  Score 203.5;  DB 4;  Length 123;
  Best Local Similarity   45.2%;  Pred. No. 1.2e-18;
  Matches   47;  Conservative    8;  Mismatches   20;  Indels   29;  Gaps    3;

Qy          1 MWLYRNPYVEAEYFPTKPMFVIAFLSPLSLIFLAKFLKKADTRDSR------QACLAASL 54
              :|||||||:|||||| ||||||||:|||||| ||||:|||     |       | | :|
Db         34 LWLYRNPYVKAEYFPTGPMFVIAFLTPLSLIFFAKFLRKADADRQRASLPRCQPCPSAKW 93

Qy         55 ALAL---------NGVFTNTIKLIVGRPRPDFFYRCFPDGLAHS 89
                |         :    :|              | | |:|||
Db         94 CLYQHHKTDSXQGHAQIAST-------------ECSPXGIAHS 123


RESULT 12
US-09-360-376-54
; Sequence 54, Application US/09360376
; Patent No. 6495739
; GENERAL INFORMATION:
;  APPLICANT:  Lassner, Michael
;  APPLICANT:  Ruezinsky, Diane
;  TITLE OF INVENTION:   PLANT PHOSPHATIDIC ACID PHOSPHATASES
;  FILE REFERENCE:   17026/01/US
;  CURRENT APPLICATION NUMBER: US/09/360,376
;  CURRENT FILING DATE:  1999-07-23
;  PRIOR APPLICATION NUMBER:   US 09/122,315
;  PRIOR FILING DATE:   1998-07-24
;  NUMBER OF SEQ ID NOS:   55
;  SOFTWARE:   PatentIn version 3.0
; SEQ ID NO 54
;   LENGTH: 282
;   TYPE: PRT
;   ORGANISM: Rattus sp.
US-09-360-376-54

  Query Match             19.1%;  Score 184.5;  DB 4;  Length 282;
  Best Local Similarity   32.4%;  Pred. No. 1.2e-15;
  Matches   60;  Conservative   17;  Mismatches   55;  Indels   53;  Gaps    8;

Qy          3 LYRNPYVEAEYFPTKPMFVIAFLSPLSLIFLAKFLKKADTRDSRQACLAASLALALNGVF 62
              |: | :|   |  |    | |||                          || : :|
Db         84 LHSNSFVSNHYIATIYKAVGAFL-------------------------FGASASQSL---- 115

Qy         63 TNTIKLIVGRPRPDFFYRCFPD--GLAHSD-----LMCTGDKDVVNEGRKSFPSGHSSFA 115
              |:  |  :|| || |   | ||   :  ||       :| |::  | ||| || ||||||:
Db        116 TDIAKYSIGRLRPHFLAVCNPDWSKINCSDGYIENFVCQGNEQKVREGRLSFYSGHSSFS 175

Qy        116 FAGLAFASFYLAGKLHCFTPQGRGK-SWRFCAFLSPLLFAAVIA------LSRTCDYKHH 168
              : | : ||      | | |  |    | |:|   ::|      |||  |||||
Db        176 MYCMLFVALYL---------QARMKGDW--ARLLRPMLQFGLVALSIYVGLSRVSDYKHH 224

Qy        169 WQDLL 173
                | |:|
Db        225 WSDVL 229


RESULT 13
US-08-992-035A-3
; Sequence 3, Application US/08992035A
; Patent No. 6242179
;  GENERAL INFORMATION:
;    APPLICANT:  Shah, Purvi
;    APPLICANT:  Hillman, Jennifer L.
;    APPLICANT:  Corley, Neil C.
;    APPLICANT:  Lal, Preeti
;    TITLE OF INVENTION:  HUMAN PHOSPHATASES
;    NUMBER OF SEQUENCES:  6
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Incyte Pharmaceuticals, Inc.
;      STREET:  3174 Porter Dr.
;      CITY:  Palo Alto
;      STATE:  CA
;      COUNTRY:  USA
;      ZIP:  94304
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Diskette
;      COMPUTER:  IBM Compatible
;      OPERATING SYSTEM:  DOS
;      SOFTWARE:  FastSEQ for Windows Version 2.0
```

```
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/992,035A
;      FILING DATE:  December 17, 1997
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:
;      FILING DATE:
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Billings, Lucy J.
;      REGISTRATION NUMBER:  36,749
;      REFERENCE/DOCKET NUMBER:  PF-0433 US
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  650-855-0555
;      TELEFAX:  650-845-4166
;  INFORMATION FOR SEQ ID NO:  3:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  283 amino acids
;      TYPE:  amino acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    IMMEDIATE SOURCE:
;      LIBRARY:  GENBANK
;      CLONE:  1487873
US-08-992-035A-3

  Query Match             19.0%;  Score 183.5;  DB 3;  Length 283;
  Best Local Similarity   31.9%;  Pred. No. 1.6e-15;
  Matches   59;  Conservative   17;  Mismatches   54;  Indels   55;  Gaps    8;

Qy          3 LYRNPYVEAEYFPTKPMFVIAFLSPLSLIFLAKFLKKADTRDSRQACLAASLALALNGVF 62
              |: | :|   | |    | |||  :|                       || :|
Db         84 LHSNSFVGNPYIATIYKAVGAFLFGVS---------------------ASQSL------ 115

Qy         63 TNTIKLIVGRPRPDFFYRCFP--------DGLAHSDLMCTGDKDVVNEGRKSFPSGHSSF 114
              |:  |  :|  || |   | |         ||    | :| |::: | ||| || ||||||
Db        116 TDIAKYTIGSLRPHFLAICNPDWSKINCSDGYI-EDYICQGNEEKVKEGRLSFYSGHSSF 174

Qy        115 AFAGLAFASFYLAGKLHCFTPQGRGK-SWRFCAFLSPLL------FAAVIALSRTCDYKH 167
              :   : | : ||         | | |  |    | |:|      |:  : |||  ||||
Db        175 SMYCMLFVALYL---------QARMKGDW--ARLLRPMLQFGLIAFSIYVGLSRVSDYKH 223

Qy        168 HWQDL 172
              || |:
Db        224 HWSDV 228


RESULT 14
US-09-360-376-53
; Sequence 53, Application US/09360376
; Patent No. 6495739
; GENERAL INFORMATION:
;  APPLICANT:  Lassner, Michael
;  APPLICANT:  Ruezinsky, Diane
;  TITLE OF INVENTION:  PLANT PHOSPHATIDIC ACID PHOSPHATASES
;  FILE REFERENCE:  17026/01/US
;  CURRENT APPLICATION NUMBER: US/09/360,376
;  CURRENT FILING DATE:  1999-07-23
;  PRIOR APPLICATION NUMBER:  US 09/122,315
;  PRIOR FILING DATE:  1998-07-24
;  NUMBER OF SEQ ID NOS:  55
;  SOFTWARE:  PatentIn version 3.0
; SEQ ID NO 53
;  LENGTH: 283
;  TYPE: PRT
;  ORGANISM: Mus musculus
US-09-360-376-53

  Query Match             19.0%;  Score 183.5;  DB 4;  Length 283;
  Best Local Similarity   31.9%;  Pred. No. 1.6e-15;
  Matches   59;  Conservative   17;  Mismatches   54;  Indels   55;  Gaps    8;

Qy          3 LYRNPYVEAEYFPTKPMFVIAFLSPLSLIFLAKFLKKADTRDSRQACLAASLALALNGVF 62
              |: | :|   | |    | |||  :|                       || :|
Db         84 LHSNSFVGNPYIATIYKAVGAFLFGVS---------------------ASQSL------ 115

Qy         63 TNTIKLIVGRPRPDFFYRCFP--------DGLAHSDLMCTGDKDVVNEGRKSFPSGHSSF 114
              |:  |  :|  || |   | |         ||    | :| |::: | ||| || ||||||
Db        116 TDIAKYTIGSLRPHFLAICNPDWSKINCSDGYI-EDYICQGNEEKVKEGRLSFYSGHSSF 174

Qy        115 AFAGLAFASFYLAGKLHCFTPQGRGK-SWRFCAFLSPLL------FAAVIALSRTCDYKH 167
              :   : | : ||         | | |  |    | |:|      |:  : |||  ||||
Db        175 SMYCMLFVALYL---------QARMKGDW--ARLLRPMLQFGLIAFSIYVGLSRVSDYKH 223

Qy        168 HWQDL 172
              || |:
Db        224 HWSDV 228


RESULT 15
US-08-992-035A-1
; Sequence 1, Application US/08992035A
; Patent No. 6242179
;  GENERAL INFORMATION:
;    APPLICANT:  Shah, Purvi
;    APPLICANT:  Hillman, Jennifer L.
;    APPLICANT:  Corley, Neil C.
;    APPLICANT:  Lal, Preeti
;    TITLE OF INVENTION:  HUMAN PHOSPHATASES
;    NUMBER OF SEQUENCES:  6
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Incyte Pharmaceuticals, Inc.
;      STREET:  3174 Porter Dr.
;      CITY:  Palo Alto
;      STATE:  CA
;      COUNTRY:  USA
;      ZIP:  94304
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Diskette
;      COMPUTER:  IBM Compatible
;      OPERATING SYSTEM:  DOS
;      SOFTWARE:  FastSEQ for Windows Version 2.0
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/992,035A
;      FILING DATE:  December 17, 1997
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:
;      FILING DATE:
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Billings, Lucy J.
;      REGISTRATION NUMBER:  36,749
;      REFERENCE/DOCKET NUMBER:  PF-0433 US
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  650-855-0555
;      TELEFAX:  650-845-4166
;  INFORMATION FOR SEQ ID NO:  1:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  285 amino acids
;      TYPE:  amino acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    IMMEDIATE SOURCE:
;      LIBRARY:  BLADNOT06
;      CLONE:  1719418
US-08-992-035A-1

  Query Match             18.0%;  Score 174.5;  DB 3;  Length 285;
  Best Local Similarity   30.9%;  Pred. No. 2.4e-14;
  Matches   60;  Conservative   25;  Mismatches   70;  Indels   39;  Gaps    9;

Qy          7 PYVEAEYFPTKPMFVIAFLSPLSLIFLAKFL--------KKADTRDSRQACLAASLALAL 58
              || ::    |  : ::    |:| | | : |          :  |::  | :  ::   |
Db         49 PYHDSTVTST-VLILVGVGLPISSIILGETLSVYCNLLHSNSFIRNNYIATIYKAIGTFL 107

Qy         59 NGV-----FTNTIKLIVGRPRPDFFYRCFPD--GLAHSD-----LMCTGDKDVVNEGRKS 106
              |       |:  |  ::|| || |   | ||   :  ||      :| |: : | ||| |
Db        108 FGAAASQSLTDIAKYSIGRLRPHFLDVCDPDWSKINCSDGYIEYYICRGNAERVKEGRLS 167
```

```
Qy      107 FPSGHSSFAFAGLAFASFYLAGKLHCFTPQGRGK-SWRFCAFLSPLLFAAVIA------L 159
            | ||||||:   : | : ||          | | |  |     | | |   ::|      |
Db      168 FYSGHSSFSMYCMLFVALYL---------QARMKGDW--ARLLRPTLQFGLVAVSIYVGL 216

Qy      160 SRTCDYKHHWQDLL 173
            ||  |||||| |:|
Db      217 SRVSDYKHHWSDVL 230
```

Search completed: August 14, 2003, 13:07:59
Job time : 31 secs

GenCore version 5.1.6

OM protein - protein search, using sw model

Run on: August 14, 2003, 13:07:30 ; Search time 56 Seconds
(without alignments)
421.082 Million cell updates/sec

Title: US-09-876-997-399
Perfect score: 968
Sequence: 1 MWLYRNPYVEAEYFPTKPMF.....RTCDYKHHWQDLLKCTNTAK 180

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 492763 seqs, 131003257 residues

Total number of hits satisfying chosen parameters: 492763

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Published_Applications_AA:*
1: /cgn2_6/ptodata/1/pubpaa/US07_PUBCOMB.pep:*
2: /cgn2_6/ptodata/1/pubpaa/PCT_NEW_PUB.pep:*
3: /cgn2_6/ptodata/1/pubpaa/US06_NEW_PUB.pep:*
4: /cgn2_6/ptodata/1/pubpaa/US06_PUBCOMB.pep:*
5: /cgn2_6/ptodata/1/pubpaa/US07_NEW_PUB.pep:*
6: /cgn2_6/ptodata/1/pubpaa/PCTUS_PUBCOMB.pep:*
7: /cgn2_6/ptodata/1/pubpaa/US08_NEW_PUB.pep:*
8: /cgn2_6/ptodata/1/pubpaa/US08_PUBCOMB.pep:*
9: /cgn2_6/ptodata/1/pubpaa/US09A_PUBCOMB.pep:*
10: /cgn2_6/ptodata/1/pubpaa/US09B_PUBCOMB.pep:*
11: /cgn2_6/ptodata/1/pubpaa/US09C_PUBCOMB.pep:*
12: /cgn2_6/ptodata/1/pubpaa/US09_NEW_PUB.pep:*
13: /cgn2_6/ptodata/1/pubpaa/US10A_PUBCOMB.pep:*
14: /cgn2_6/ptodata/1/pubpaa/US10B_PUBCOMB.pep:*
15: /cgn2_6/ptodata/1/pubpaa/US10C_PUBCOMB.pep:*
16: /cgn2_6/ptodata/1/pubpaa/US10_NEW_PUB.pep:*
17: /cgn2_6/ptodata/1/pubpaa/US60_NEW_PUB.pep:*
18: /cgn2_6/ptodata/1/pubpaa/US60_PUBCOMB.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	968	100.0	180	10	US-09-731-872-399	Sequence 399, App
2	641	66.2	233	15	US-10-103-313-562	Sequence 562, App
3	633	65.4	235	12	US-10-241-220-99	Sequence 99, Appl
4	389	40.2	128	11	US-09-866-050A-316	Sequence 316, App
5	329	34.0	149	15	US-10-103-313-452	Sequence 452, App
6	203.5	21.0	123	11	US-09-866-050A-170	Sequence 170, App
7	138	14.3	220	15	US-10-106-698-5750	Sequence 5750, Ap
8	115	11.9	180	10	US-09-860-670-125	Sequence 125, App
9	115	11.9	180	15	US-10-103-313-490	Sequence 490, App
10	115	11.9	183	15	US-10-103-313-340	Sequence 340, App
11	110	11.4	203	15	US-10-103-313-493	Sequence 493, App
12	93	9.6	525	15	US-10-156-761-15017	Sequence 15017, A
13	87	9.0	84	9	US-09-864-761-40341	Sequence 40341, A
14	82	8.5	497	15	US-10-156-761-9418	Sequence 9418, Ap
15	77.5	8.0	184	9	US-09-925-299-860	Sequence 860, App
16	77.5	8.0	184	11	US-09-925-299-860	Sequence 860, App
17	73.5	7.6	283	15	US-10-156-761-8708	Sequence 8708, Ap
18	73	7.5	216	9	US-09-815-242-13375	Sequence 13375, A
19	72	7.4	580	15	US-10-004-551-10	Sequence 10, Appl
20	70.5	7.3	159	10	US-09-738-626-6679	Sequence 6679, Ap
21	68	7.0	276	9	US-09-741-669-302	Sequence 302, App
22	67.5	7.0	444	9	US-09-853-386-132	Sequence 132, App
23	67.5	7.0	444	9	US-09-853-386-136	Sequence 136, App
24	67.5	7.0	444	9	US-09-191-724-15	Sequence 15, Appl
25	67.5	7.0	444	15	US-10-242-822B-2	Sequence 2, Appli
26	67	6.9	410	15	US-10-156-761-13608	Sequence 13608, A
27	66.5	6.9	227	8	US-08-834-705-16	Sequence 16, Appl
28	66.5	6.9	227	9	US-09-815-242-11375	Sequence 11375, A
29	66.5	6.9	228	9	US-09-815-242-11542	Sequence 11542, A
30	66.5	6.9	274	15	US-10-156-761-10121	Sequence 10121, A
31	66.5	6.9	705	11	US-09-907-907A-43	Sequence 43, Appl
32	65.5	6.8	296	10	US-09-771-730-69	Sequence 69, Appl
33	65.5	6.8	296	10	US-09-771-730-116	Sequence 116, App
34	65.5	6.8	296	10	US-09-771-730-120	Sequence 120, App
35	65.5	6.8	303	10	US-09-771-730-118	Sequence 118, App
36	65.5	6.8	332	10	US-09-886-055-63	Sequence 63, Appl
37	65.5	6.8	332	10	US-09-771-730-10	Sequence 10, Appl
38	65.5	6.8	332	10	US-09-771-730-28	Sequence 28, Appl
39	65.5	6.8	332	11	US-09-804-291-63	Sequence 63, Appl
40	65.5	6.8	332	12	US-10-017-161-244	Sequence 244, App
41	65.5	6.8	332	15	US-10-236-250-2	Sequence 2, Appli
42	65.5	6.8	380	9	US-09-815-242-11291	Sequence 11291, A
43	65.5	6.8	383	12	US-10-259-165-202	Sequence 202, App
44	65.5	6.8	453	15	US-10-220-382-19	Sequence 19, Appl
45	65	6.7	320	15	US-10-242-822B-32	Sequence 32, Appl

ALIGNMENTS

RESULT 1
US-09-731-872-399
; Sequence 399, Application US/09731872
; Patent No. US20020102604A1
; GENERAL INFORMATION:
; APPLICANT: Dumas Milne Edwards, Jean Baptiste
; APPLICANT: Bougueleret, Lydie
; APPLICANT: Jobert, Severin
; TITLE OF INVENTION: FULL-LENGTH HUMAN cDNAs ENCODING POTENTIALLY SECRETED PROTEINS
; FILE REFERENCE: 78.US3.REG
; CURRENT APPLICATION NUMBER: US/09/731,872
; CURRENT FILING DATE: 2000-12-07
; PRIOR APPLICATION NUMBER: US 60/169,629
; PRIOR FILING DATE: 1999-12-08
; PRIOR APPLICATION NUMBER: US 60/187,470
; PRIOR FILING DATE: 2000-03-06
; NUMBER OF SEQ ID NOS: 482
; SOFTWARE: Patent.pm
; SEQ ID NO 399
; LENGTH: 180
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-731-872-399

Query Match 100.0%; Score 968; DB 10; Length 180;
Best Local Similarity 100.0%; Pred. No. 1.3e-108;
Matches 180; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy          1 MWLYRNPYVEAEYFPTKPMFVIAFLSPLSLIFLAKFLKKADTRDSRQACLAASLALALNG 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MWLYRNPYVEAEYFPTKPMFVIAFLSPLSLIFLAKFLKKADTRDSRQACLAASLALALNG 60

Qy         61 VFTNTIKLIVGRPRPDFFYRCFPDGLAHSDLMCTGDKDVVNEGRKSFPSGHSSFAFAGLA 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 VFTNTIKLIVGRPRPDFFYRCFPDGLAHSDLMCTGDKDVVNEGRKSFPSGHSSFAFAGLA 120

Qy        121 FASFYLAGKLHCFTPQGRGKSWRFCAFLSPLLFAAVIALSRTCDYKHHWQDLLKCTNTAK 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
Db        121 FASFYLAGKLHCFTPQGRGKSWRFCAFLSPLLFAAVIALSRTCDYKHHWQDLLKCTNTAK 180


RESULT 2
US-10-103-313-562
; Sequence 562, Application US/10103313
; Publication No. US20030082758A1
; GENERAL INFORMATION:
;  APPLICANT: Rosen et al.
;  TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
;  FILE REFERENCE: PJZ07C1
;  CURRENT APPLICATION NUMBER: US/10/103,313
;  CURRENT FILING DATE:  2002-03-12
;  NUMBER OF SEQ ID NOS: 653
;       Prior Application removed - See File Wrapper or Palm
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 562
;   LENGTH: 233
;   TYPE: PRT
;   ORGANISM: Homo sapiens
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (14)
;   OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
US-10-103-313-562

  Query Match             66.2%;  Score 641;  DB 15;  Length 233;
  Best Local Similarity   66.7%;  Pred. No. 5.1e-69;
  Matches  114;  Conservative   27;  Mismatches   30;  Indels    0;  Gaps    0;

Qy          1 MWLYRNPYVEAEYFPTKPMFVIAFLSPLSLIFLAKFLKKADTRDSRQACLAASLALALNG 60
              :|||:|| |:::  ||: || |:||:||::| : | ::: |  : ::| || ||||||||
Db         47 IWLYKNPLVQSDNIPTRLMFAISFLTPLAVICVVKIIRRTDKTEIKEAFLAVSLALALNG 106

Qy         61 VFTNTIKLIVGRPRPDFFYRCFPDGLAHSDLMCTGDKDVVNEGRKSFPSGHSSFAFAGLA 120
              | |||||||||||||||||||||||: :|:: |||| |:|:|||||||| |||||:||
Db        107 VCTNTIKLIVGRPRPDFFYRCFPDGVMNSEMHCTGDPDLVSEGRKSFPSIHSSFAFSGLG 166

Qy        121 FASFYLAGKLHCFTPQGRGKSWRFCAFLSPLLFAAVIALSRTCDYKHHWQD 171
              | :|||||||||||  ||||||| || : ||  | :||||| |||||||||
Db        167 FTTFYLAGKLHCFTESGRGKSWRLCAAILPLYCAMMIALSRMCDYKHHWQD 217


RESULT 3
US-10-241-220-99
; Sequence 99, Application US/10241220
; Publication No. US20030148408A1
; GENERAL INFORMATION:
;  APPLICANT: Frantz,Gretchen
;  APPLICANT:  Hillan,Kenneth J.
;  APPLICANT:  Phillips,Heidi
;  APPLICANT:  Polakis,Paul
;  APPLICANT:  Spencer,Susan
;  APPLICANT:  Williams,P.Mickey
;  APPLICANT:  Wu,Thomas
;  APPLICANT:  Zhang,Zemin
;  TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE DIAGNOSIS AND
;  TITLE OF INVENTION:  TREATMENT OF TUMOR
;  FILE REFERENCE: P5010R1-US
;  CURRENT APPLICATION NUMBER: US/10/241,220
;  CURRENT FILING DATE:  2002-12-13
;  NUMBER OF SEQ ID NOS: 120
; SEQ ID NO 99
;   LENGTH: 235
;   TYPE: PRT
;   ORGANISM: Homo Sapien
US-10-241-220-99

  Query Match             65.4%;  Score 633;  DB 12;  Length 235;
  Best Local Similarity   66.1%;  Pred. No. 4.8e-68;
  Matches  113;  Conservative   27;  Mismatches   31;  Indels    0;  Gaps    0;

Qy          1 MWLYRNPYVEAEYFPTKPMFVIAFLSPLSLIFLAKFLKKADTRDSRQACLAASLALALNG 60
              :|||:|| |:::  ||: || |:||:||::| : | ::: |  : ::| || ||||||||
Db         36 IWLYKNPLVQSDNIPTRLMFAISFLTPLAVICVVKIIRRTDKTEIKEAFLAVSLALALNG 95

Qy         61 VFTNTIKLIVGRPRPDFFYRCFPDGLAHSDLMCTGDKDVVNEGRKSFPSGHSSFAFAGLA 120
              | |||||||||||| |||||||||: :|:: |||| |:|:|||||||| |||||:||
Db         96 VCTNTIKLIVGRPRADFFYRCFPDGVMNSEMHCTGDPDLVSEGRKSFPSIHSSFAFSGLG 155

Qy        121 FASFYLAGKLHCFTPQGRGKSWRFCAFLSPLLFAAVIALSRTCDYKHHWQD 171
              | :|||||||||||  ||||||| || : ||  | :||||| |||||||||
Db        156 FTTFYLAGKLHCFTESGRGKSWRLCAAILPLYCAMMIALSRMCDYKHHWQD 206


RESULT 4
US-09-866-050A-316
; Sequence 316, Application US/09866050A
; Publication No. US20030040471A1
; GENERAL INFORMATION:
;  APPLICANT: Watson, James D.
;  APPLICANT:  Strachan, Lorna
;  APPLICANT:  Sleeman, Matthew
;  APPLICANT:  Onrust, Rene
;  APPLICANT:  Murison, James G.
;  APPLICANT:  Kumble, Krishanand D.
;  TITLE OF INVENTION: Compositions Isolated From Skin Cells
;  TITLE OF INVENTION:  and Methods for Their Use
;  FILE REFERENCE: 11000.1011c4U
;  CURRENT APPLICATION NUMBER: US/09/866,050A
;  CURRENT FILING DATE:  2001-05-24
;  NUMBER OF SEQ ID NOS: 725
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 316
;   LENGTH: 128
;   TYPE: PRT
;   ORGANISM: Rat
US-09-866-050A-316

  Query Match             40.2%;  Score 389;  DB 11;  Length 128;
  Best Local Similarity   89.0%;  Pred. No. 5.8e-39;
  Matches   73;  Conservative    4;  Mismatches    5;  Indels    0;  Gaps    0;

Qy          1 MWLYRNPYVEAEYFPTKPMFVIAFLSPLSLIFLAKFLKKADTRDSRQACLAASLALALNG 60
              :|||||||||||||| |||||||:|||||| ||||:|||  ||:|||||||||||||
Db         47 LWLYRNPYVEAEYFPTGPMFVIAFLTPLSLIFFAKFLRKADATDSKQACLAASLALALNG 106

Qy         61 VFTNTIKLIVGRPRPDFFYRCF 82
              |||| ||||||||||||||||||
Db        107 VFTNIIKLIVGRPRPDFFYRCF 128


RESULT 5
US-10-103-313-452
; Sequence 452, Application US/10103313
; Publication No. US20030082758A1
; GENERAL INFORMATION:
;  APPLICANT: Rosen et al.
;  TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
;  FILE REFERENCE: PJZ07C1
;  CURRENT APPLICATION NUMBER: US/10/103,313
;  CURRENT FILING DATE:  2002-03-12
;  NUMBER OF SEQ ID NOS: 653
;       Prior Application removed - See File Wrapper or Palm
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 452
;   LENGTH: 149
;   TYPE: PRT
;   ORGANISM: Homo sapiens
US-10-103-313-452

  Query Match             34.0%;  Score 329;  DB 15;  Length 149;
  Best Local Similarity   75.6%;  Pred. No. 1.2e-31;
  Matches   59;  Conservative    6;  Mismatches   13;  Indels    0;  Gaps    0;

Qy         94 TGDKDVVNEGRKSFPSGHSSFAFAGLAFASFYLAGKLHCFTPQGRGKSWRFCAFLSPLLF 153
```

```
              ||| |:|:|||||||| ||||||:|| | :|||||||||||  ||||||| || : ||
Db          7 TGDPDLVSEGRKSFPSIHSSFAFSGLGFTTFYLAGKLHCFTESGRGKSWRLCAAILPLYC 66

Qy        154 AAVIALSRTCDYKHHWQD 171
              | :||||| |||||||||
Db         67 AMMIALSRMCDYKHHWQD 84


RESULT 6
US-09-866-050A-170
; Sequence 170, Application US/09866050A
; Publication No. US20030040471A1
; GENERAL INFORMATION:
;  APPLICANT: Watson, James D.
;  APPLICANT:  Strachan, Lorna
;  APPLICANT:  Sleeman, Matthew
;  APPLICANT:  Onrust, Rene
;  APPLICANT:  Murison, James G.
;  APPLICANT:  Kumble, Krishanand D.
;  TITLE OF INVENTION: Compositions Isolated From Skin Cells
;  TITLE OF INVENTION:  and Methods for Their Use
;  FILE REFERENCE: 11000.1011c4U
;  CURRENT APPLICATION NUMBER: US/09/866,050A
;  CURRENT FILING DATE:  2001-05-24
;  NUMBER OF SEQ ID NOS: 725
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 170
;   LENGTH: 123
;   TYPE: PRT
;   ORGANISM: Rat
;   FEATURE:
;   NAME/KEY: VARIANT
;   LOCATION: (1)...(123)
;   OTHER INFORMATION: Xaa = Any Amino Acid
US-09-866-050A-170

  Query Match             21.0%;  Score 203.5;  DB 11;  Length 123;
  Best Local Similarity   45.2%;  Pred. No. 1.3e-16;
  Matches   47;  Conservative    8;  Mismatches   20;  Indels   29;  Gaps    3;

Qy          1 MWLYRNPYVEAEYFPTKPMFVIAFLSPLSLIFLAKFLKKADTRDSR------QACLAASL 54
              :|||||||:|||||| ||||||||:|||||| ||||:|||    |      | | :|
Db         34 LWLYRNPYVKAEYFPTGPMFVIAFLTPLSLIFFAKFLRKADADRQRASLPRCQPCPSAKW 93

Qy         55 ALAL---------NGVFTNTIKLIVGRPRPDFFYRCFPDGLAHS 89
                |          :    :|              | | |:|||
Db         94 CLYQHHKTDSXQGHAQIAST-------------ECSPXGIAHS 123


RESULT 7
US-10-106-698-5750
; Sequence 5750, Application US/10106698
; Publication No. US20030109690A1
; GENERAL INFORMATION:
;  APPLICANT: Ruben et al.
;  TITLE OF INVENTION: Colon and Colon Cancer Associated Polynucleotides and Polypeptide
;  FILE REFERENCE: PA005P1
;  CURRENT APPLICATION NUMBER: US/10/106,698
;  CURRENT FILING DATE:  2002-03-27
;  PRIOR APPLICATION NUMBER: PCT/US00/26524
;  PRIOR FILING DATE: 2000-09-28
;  PRIOR APPLICATION NUMBER: US 60/157,137
;  PRIOR FILING DATE: 1999-09-29
;  PRIOR APPLICATION NUMBER: US 60/163,280
;  PRIOR FILING DATE: 1999-11-03
;  NUMBER OF SEQ ID NOS: 8564
;  SOFTWARE: PatentIn Ver. 3.0
; SEQ ID NO 5750
;   LENGTH: 220
;   TYPE: PRT
;   ORGANISM: Homo sapiens
;   FEATURE:
;   NAME/KEY: MISC_FEATURE
;   LOCATION: (65)
;   OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
;   NAME/KEY: MISC_FEATURE
;   LOCATION: (96)
;   OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
;   NAME/KEY: MISC_FEATURE
;   LOCATION: (117)
;   OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
US-10-106-698-5750

  Query Match             14.3%;  Score 138;  DB 15;  Length 220;
  Best Local Similarity   26.7%;  Pred. No. 2.3e-08;
  Matches   35;  Conservative   24;  Mismatches   58;  Indels   14;  Gaps    4;

Qy         53 SLALALNGVFTNTIKLIVGRPRPDFFYRCFP-----DGLAHSDL-----MCTGDKDVVNE 102
              :  |    :| |  ::: |   | |   | |     |  ||        :||||::|: :
Db         27 AFGLFATDIFVNAGQVVTGHLTPYFLTVCKPNYTSADCXAHHQFINNGNICTGDREVIEK 86

Qy        103 GRKSFPSGHSSFAFAGLAFASFYLAGKLHCFTPQGRGKSWRFCAFLSPLLFAAVIALSRT 162
              |:|||| | : :     :|: |:   :   |  |      |  |  |  | :  |:|
Db         87 ARRSFPSKHXALSIYSALYATMYITSTIK--TXSSRLAKPVLC--LGTLCTAFLTGLNRV 142

Qy        163 CDYKHHWQDLL 173
               :|::|  |::
Db        143 SEYRNHCSDVI 153


RESULT 8
US-09-860-670-125
; Sequence 125, Application US/09860670
; Patent No. US20020165137A1
; GENERAL INFORMATION:
;  APPLICANT: Ruben et al.
;  TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
;  FILE REFERENCE: PA127P1
;  CURRENT APPLICATION NUMBER: US/09/860,670
;  CURRENT FILING DATE:  2001-05-21
;  Prior application data removed - consult PALM or file wrapper
;  NUMBER OF SEQ ID NOS: 289
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 125
;   LENGTH: 180
;   TYPE: PRT
;   ORGANISM: Homo sapiens
;   FEATURE:
;   NAME/KEY: SITE
;   LOCATION: (154)
;   OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
;   NAME/KEY: SITE
;   LOCATION: (155)
;   OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
;   NAME/KEY: SITE
;   LOCATION: (169)
;   OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
;   NAME/KEY: SITE
;   LOCATION: (175)
;   OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
US-09-860-670-125

  Query Match             11.9%;  Score 115;  DB 10;  Length 180;
  Best Local Similarity   23.2%;  Pred. No. 1e-05;
  Matches   29;  Conservative   24;  Mismatches   48;  Indels   24;  Gaps    4;

Qy         64 NTIKLIVGRPRPDFFYRCFPDGLA----------HSDLMCTGDKDVVNEGRKSFPSGHSS 113
              |  ::: |   | |   | |: |            :  |||: |::   ||:|||  ::
Db          2 NAGQVVTGNLAPHFLALCKPNYTALGCQQYTQFISGEEACTGNPDLIMRARKTFPSKEAA 61

Qy        114 FAFAGLAFASFYLAGKLHCFTPQGRGKSWR-----FCAFLSPLLFAAVIALSRTCDYKHH 168
               :     : : |:   :       : | |      |  |  | |  :  |:|  :|::|
Db         62 LSVYAAMYLTMYITNTI-------KAKGTRLAKPVLCLGLMCLAF--LTGLNRVAEYRNH 112

Qy        169 WQDLL 173
               | |::
```

```
Db        113 WSDVI 117


RESULT 9
US-10-103-313-490
; Sequence 490, Application US/10103313
; Publication No. US20030082758A1
; GENERAL INFORMATION:
;  APPLICANT: Rosen et al.
;  TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
;  FILE REFERENCE: PJZ07C1
;  CURRENT APPLICATION NUMBER: US/10/103,313
;  CURRENT FILING DATE:  2002-03-12
;  NUMBER OF SEQ ID NOS: 653
;      Prior Application removed - See File Wrapper or Palm
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 490
;   LENGTH: 180
;   TYPE: PRT
;   ORGANISM: Homo sapiens
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (154)
;   OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
;   NAME/KEY: misc_feature
;   LOCATION: (155)
;   OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
;   NAME/KEY: misc_feature
;   LOCATION: (169)
;   OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
;   NAME/KEY: misc_feature
;   LOCATION: (175)
;   OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
US-10-103-313-490

  Query Match             11.9%;  Score 115;  DB 15;  Length 180;
  Best Local Similarity   23.2%;  Pred. No. 1e-05;
  Matches   29;  Conservative   24;  Mismatches   48;  Indels   24;  Gaps    4;

Qy         64 NTIKLIVGRPRPDFFYRCFPDGLA----------HSDLMCTGDKDVVNEGRKSFPSGHSS 113
              |  ::: |   | |   | |: |            :  |||: |::   ||:||| ::
Db          2 NAGQVVTGNLAPHFLALCKPNYTALGCQQYTQFISGEEACTGNPDLIMRARKTFPSKEAA 61

Qy        114 FAFAGLAFASFYLAGKLHCFTPQGRGKSWR-----FCAFLSPLLFAAVIALSRTCDYKHH 168
                :     : : |:   :       : |   |     | | | |  :  |:|  :|::|
Db         62 LSVYAAMYLTMYITNTI-------KAKGTRLAKPVLCLGLMCLAF--LTGLNRVAEYRNH 112

Qy        169 WQDLL 173
                | |::
Db        113 WSDVI 117


RESULT 10
US-10-103-313-340
; Sequence 340, Application US/10103313
; Publication No. US20030082758A1
; GENERAL INFORMATION:
;  APPLICANT: Rosen et al.
;  TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
;  FILE REFERENCE: PJZ07C1
;  CURRENT APPLICATION NUMBER: US/10/103,313
;  CURRENT FILING DATE:  2002-03-12
;  NUMBER OF SEQ ID NOS: 653
;      Prior Application removed - See File Wrapper or Palm
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 340
;   LENGTH: 183
;   TYPE: PRT
;   ORGANISM: Homo sapiens
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (172)
;   OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
;   NAME/KEY: misc_feature
;   LOCATION: (178)
;   OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
US-10-103-313-340

  Query Match             11.9%;  Score 115;  DB 15;  Length 183;
  Best Local Similarity   23.2%;  Pred. No. 1.1e-05;
  Matches   29;  Conservative   24;  Mismatches   48;  Indels   24;  Gaps    4;

Qy         64 NTIKLIVGRPRPDFFYRCFPDGLA----------HSDLMCTGDKDVVNEGRKSFPSGHSS 113
              |  ::: |   | |   | |: |            :  |||: |::   ||:||| ::
Db          5 NAGQVVTGNLAPHFLALCKPNYTALGCQQYTQFISGEEACTGNPDLIMRARKTFPSKEAA 64

Qy        114 FAFAGLAFASFYLAGKLHCFTPQGRGKSWR-----FCAFLSPLLFAAVIALSRTCDYKHH 168
                :     : : |:   :       : |   |     | | | |  :  |:|  :|::|
Db         65 LSVYAAMYLTMYITNTI-------KAKGTRLAKPVLCLGLMCLAF--LTGLNRVAEYRNH 115

Qy        169 WQDLL 173
                | |::
Db        116 WSDVI 120


RESULT 11
US-10-103-313-493
; Sequence 493, Application US/10103313
; Publication No. US20030082758A1
; GENERAL INFORMATION:
;  APPLICANT: Rosen et al.
;  TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
;  FILE REFERENCE: PJZ07C1
;  CURRENT APPLICATION NUMBER: US/10/103,313
;  CURRENT FILING DATE:  2002-03-12
;  NUMBER OF SEQ ID NOS: 653
;      Prior Application removed - See File Wrapper or Palm
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 493
;   LENGTH: 203
;   TYPE: PRT
;   ORGANISM: Homo sapiens
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (73)
;   OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
;   NAME/KEY: misc_feature
;   LOCATION: (186)
;   OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
;   NAME/KEY: misc_feature
;   LOCATION: (190)
;   OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
;   NAME/KEY: misc_feature
;   LOCATION: (196)
;   OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
US-10-103-313-493

  Query Match             11.4%;  Score 110;  DB 15;  Length 203;
  Best Local Similarity   23.2%;  Pred. No. 4.8e-05;
  Matches   33;  Conservative   18;  Mismatches   53;  Indels   38;  Gaps    5;

Qy         60 GVFTNTI-----KLIVGRPRPDFFYRC----------------FPDGLAHSDLMCTGDK 97
              |:|| ||      ::: | | | |  |                ||       | |
Db         29 GLFTTTIFANAGQVVTGNPTPHFLSVCRPNYTALGCLPPSPDRPXPDRFVTDQGACAGSP 88

Qy         98 DVVNEGRKSFPSGHSSFAFAGLAFASFYLAGKLHCFTPQGRGKSWRF------CAFLSPL 151
              :|   |::||   ::     : : : |:      |  | | |       | | |
Db         89 SLVAAARRAFPCKDAALCAYAVTYTAMYV-------TLVFRVKGSRLVKPSLCLALLCPA 141

Qy        152 LFAAVIALSRTCDYKHHWQDLL 173
                 |:   |  :|::|| |:|
Db        142 FLVGVV---RVAEYRNHWSDVL 160


RESULT 12
US-10-156-761-15017
```

; Sequence 15017, Application US/10156761
; Publication No. US20030119018A1
; GENERAL INFORMATION:
; APPLICANT: OMURA, SATOSHI
; APPLICANT: IKEDA, HARUO
; APPLICANT: ISHIKAWA, JUN
; APPLICANT: HORIKAWA, HIROSHI
; APPLICANT: SHIBA, TADAYOSHI
; APPLICANT: SAKAKI, YOSHIYUKI
; APPLICANT: HATTORI, MASAHIRA
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
; FILE REFERENCE: 249-262
; CURRENT APPLICATION NUMBER: US/10/156,761
; CURRENT FILING DATE: 2002-05-29
; PRIOR APPLICATION NUMBER: JP 2001-204089
; PRIOR FILING DATE: 2001-05-30
; PRIOR APPLICATION NUMBER: JP 2001-272697
; PRIOR FILING DATE: 2001-08-02
; NUMBER OF SEQ ID NOS: 15109
; SEQ ID NO 15017
; LENGTH: 525
; TYPE: PRT
; ORGANISM: Streptomyces avermitilis
US-10-156-761-15017

Query Match 9.6%; Score 93; DB 15; Length 525;
Best Local Similarity 30.3%; Pred. No. 0.019;
Matches 46; Conservative 16; Mismatches 38; Indels 52; Gaps 8;

```
Qy          34 AKFLKKADTRDSRQACLAASLALALNGVFTNTI-KLIVGRPRPDFFYRCFPDGLAHSDLM 92
               |  |  || | :|:| :    :|||  :  ||: |    ||||               |:
Db          44 AAVLATADGRTARRAAVRGVGSLALASLTVNTVAKWATRRPRP--------------LL 88

Qy          93 CTGDKDVVNEGR--------KSFPSGHSSFAFA---GLAFASFYLAGKLHCFTPQGRGKS 141
                  |:| : |         |||||||: | |   |:|  |              | |
Db          89 -----DLVPQIRHLTRQPHTTSFPSGHSASAAAFATGVALES-----------TGYG-- 129

Qy         142 WRFCAFLSPLLFAAVIALSRTCDYKHHWQDLL 173
                   | ::||  || :| ||     |:  |:|
Db         130 ----ALVAPL--AAAVAFSRVYVGVHYPGDVL 155
```

RESULT 13
US-09-864-761-40341
; Sequence 40341, Application US/09864761
; Patent No. US20020048763A1
; GENERAL INFORMATION:
; APPLICANT: Penn, Sharron G.
; APPLICANT: Rank, David R.
; APPLICANT: Hanzel, David K.
; APPLICANT: Chen, Wensheng
; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR
; TITLE OF INVENTION: GENE EXPRESSION ANALYSIS BY MICROARRAY
; FILE REFERENCE: Aeomica-X-1
; CURRENT APPLICATION NUMBER: US/09/864,761
; CURRENT FILING DATE: 2001-05-23
; PRIOR APPLICATION NUMBER: US 60/180,312
; PRIOR FILING DATE: 2000-02-04
; PRIOR APPLICATION NUMBER: US 60/207,456
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: US 09/632,366
; PRIOR FILING DATE: 2000-08-03
; PRIOR APPLICATION NUMBER: GB 24263.6
; PRIOR FILING DATE: 2000-10-04
; PRIOR APPLICATION NUMBER: US 60/236,359
; PRIOR FILING DATE: 2000-09-27
; PRIOR APPLICATION NUMBER: PCT/US01/00666
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00667
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00664
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00669
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00665
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00668
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00663
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00662
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00661
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00670
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: US 60/234,687
; PRIOR FILING DATE: 2000-09-21
; PRIOR APPLICATION NUMBER: US 09/608,408
; PRIOR FILING DATE: 2000-06-30
; PRIOR APPLICATION NUMBER: US 09/774,203
; PRIOR FILING DATE: 2001-01-29
; NUMBER OF SEQ ID NOS: 49117
; SOFTWARE: Annomax Sequence Listing Engine vers. 1.1
; SEQ ID NO 40341
; LENGTH: 84
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: MAP TO AL161744.1
; OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 14
; OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 13
; OTHER INFORMATION: EXPRESSED IN HEART, SIGNAL = 14
; OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 11
; OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 16
; OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 10
; OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 12
; OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 14
; OTHER INFORMATION: SWISSPROT HIT: Q10022, EVALUE 1.00e-05
; OTHER INFORMATION: EST_HUMAN HIT: AW157233.1, EVALUE 3.00e-43
US-09-864-761-40341

Query Match 9.0%; Score 87; DB 9; Length 84;
Best Local Similarity 25.0%; Pred. No. 0.009;
Matches 21; Conservative 16; Mismatches 33; Indels 14; Gaps 2;

```
Qy          53 SLALALNGVFTNTIKLIVGRPRPDFFYRCFPDGLA----------HSDLMCTGDKDVVNE 102
               :  |    :| |  ::: |   | |   | |:  |            :  |||: |::
Db           3 TFGLFATDIFVNAGQVVTGNLAPHFLALCKPNYTALGCQQYTQFISGEEACTGNPDLIMR 62

Qy         103 GRKSFPSGHSSFAFAGLAFASFYL 126
                ||:|||  :: :     :|: ||
Db          63 ARKTFPSKEAALS----VYAAMYL 82
```

RESULT 14
US-10-156-761-9418
; Sequence 9418, Application US/10156761
; Publication No. US20030119018A1
; GENERAL INFORMATION:
; APPLICANT: OMURA, SATOSHI
; APPLICANT: IKEDA, HARUO
; APPLICANT: ISHIKAWA, JUN
; APPLICANT: HORIKAWA, HIROSHI
; APPLICANT: SHIBA, TADAYOSHI
; APPLICANT: SAKAKI, YOSHIYUKI
; APPLICANT: HATTORI, MASAHIRA
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
; FILE REFERENCE: 249-262
; CURRENT APPLICATION NUMBER: US/10/156,761
; CURRENT FILING DATE: 2002-05-29
; PRIOR APPLICATION NUMBER: JP 2001-204089
; PRIOR FILING DATE: 2001-05-30
; PRIOR APPLICATION NUMBER: JP 2001-272697
; PRIOR FILING DATE: 2001-08-02
; NUMBER OF SEQ ID NOS: 15109

```
; SEQ ID NO 9418
;   LENGTH: 497
;   TYPE: PRT
;   ORGANISM: Streptomyces avermitilis
US-10-156-761-9418

  Query Match             8.5%;  Score 82;  DB 15;  Length 497;
  Best Local Similarity   29.8%;  Pred. No. 0.37;
  Matches   42;  Conservative   11;  Mismatches   36;  Indels   52;  Gaps    7;

Qy          45 SRQACLAASLALALNGVFTNTI-KLIVGRPRPDFFYRCFPDGLAHSDLMCTGDKDVVNEG 103
               :|:|      :| |     ||: |  | ||||                 | |
Db          72 ARRAAARGLASLTLASATINTLGKRSVRRPRPVL-------------------DPVPHV 111

Qy         104 RK--------SFPSGHSSFAFA---GLAFASFYLAGKLHCFTPQGRGKSWRFCAFLSPLL 152
               |:        |||||||: | |   |:|  |            :| |      | |:||
Db         112 RRLKRQPITTSFPSGHSASAAAFATGVALES-----------RGLG------AALAPL- 152

Qy         153 FAAVIALSRTCDYKHHWQDLL 173
                || :||||     |   |:|
Db         153 -AAAVALSRVYTGVHFPSDVL 172


RESULT 15
US-09-925-299-860
; Sequence 860, Application US/09925299
; Patent No. US20020055627A1
; GENERAL INFORMATION:
;  APPLICANT: Rosen et al.
;  TITLE OF INVENTION: Nucleic Acids, Proteins and Antibodies
;  FILE REFERENCE: PA102
;  CURRENT APPLICATION NUMBER: US/09/925,299
;  CURRENT FILING DATE:  2001-08-10
;  PRIOR APPLICATION NUMBER: PCT/US00/05883
;  PRIOR FILING DATE: 2000-03-08
;  PRIOR APPLICATION NUMBER: 60/124,270
;  PRIOR FILING DATE: 1999-03-12
;  NUMBER OF SEQ ID NOS: 1556
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 860
;   LENGTH: 184
;   TYPE: PRT
;   ORGANISM: Homo sapiens
;   FEATURE:
;   NAME/KEY: SITE
;   LOCATION: (174)
;   OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
US-09-925-299-860

  Query Match             8.0%;  Score 77.5;  DB 9;  Length 184;
  Best Local Similarity   25.8%;  Pred. No. 0.35;
  Matches   31;  Conservative   13;  Mismatches   55;  Indels   21;  Gaps    4;

Qy          56 LALNGVFTNTIKLIVGRPRPDFFYRCFPDGLAHSDLMCTGDKDVVNEGRKSFPSGHSSFA 115
               |||.|      || ::  |||               | |   |   :   || || |
Db          12 LALNEGVNWLIKNVIQEPRP----------------CGGPHTAVGT-KYGMPSSHSQFM 53

Qy         116 FAGLAFASFYLAGKLHCFTPQGR--GKSWRFCAFLSPLLFAAVIALSRTCDYKHHWQDLL 173
               :    ::  :|  ::|  |   |     ||    |  |  | ::: ||     | |  :|
Db          54 WFFSVYSFLFLYLRMH-QTNNARFLDLLWRHVLSLGLLAVAFLVSYSRVYLLYHTWSQVL 112


Search completed: August 14, 2003, 13:16:39
Job time : 57 secs
```

GenCore version 5.1.6

OM protein - protein search, using sw model

Run on: August 14, 2003, 13:03:37 ; Search time 39 Seconds
(without alignments)
443.855 Million cell updates/sec

Title: US-09-876-997-399
Perfect score: 968
Sequence: 1 MWLYRNPYVEAEYFPTKPMF.....RTCDYKHHWQDLLKCTNTAK 180

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283308 seqs, 96168682 residues

Total number of hits satisfying chosen parameters: 283308

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : PIR_76:*
1: pir1:*
2: pir2:*
3: pir3:*
4: pir4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	372	38.4	290	2	E86284	F9L1.2 protein - A
2	337	34.8	302	2	E84421	probable phosphati
3	247	25.5	289	2	S70114	probable membrane
4	242	25.0	279	2	T40445	phosphatidic acid
5	200.5	20.7	274	2	S69561	hypothetical prote
6	143	14.8	341	2	T16951	hypothetical prote
7	124	12.8	239	2	S64327	probable membrane
8	118	12.2	318	2	JE0199	phosphatidic acid
9	118	12.2	318	2	H89632	protein F13E6.5 [i
10	115.5	11.9	186	1	D69206	conserved hypothet
11	112	11.6	218	2	E86750	hypothetical prote
12	104	10.7	217	2	D97672	hypothetical prote
13	104	10.7	242	2	AB2897	conserved hypothet
14	102.5	10.6	396	2	T24576	hypothetical prote
15	97.5	10.1	292	2	AF3647	phosphatidylglycer
16	93.5	9.7	213	2	T05449	hypothetical prote
17	93.5	9.7	217	2	AG1277	hypothetical prote
18	92	9.5	325	2	S66668	hydrogen peroxide-
19	90.5	9.3	330	1	F64346	hypothetical prote
20	89	9.2	259	2	A87623	PAP2 homolog prote
21	87	9.0	1294	2	T19982	hypothetical prote
22	85.5	8.8	217	2	AG1640	hypothetical prote
23	83	8.6	322	2	T45568	hypothetical prote
24	78.5	8.1	238	2	E70524	hypothetical prote
25	78.5	8.1	260	2	D75404	phosphatidylglycer
26	77.5	8.0	378	2	G70435	hypothetical prote
27	77	8.0	184	2	A71046	hypothetical prote
28	77	8.0	274	2	A82509	probable phosphati
29	77	8.0	298	2	T13684	NADH2 dehydrogenas
30	76.5	7.9	232	2	AD2542	hypothetical prote
31	76.5	7.9	649	2	D90496	hypothetical prote
32	76.5	7.9	838	2	S28911	gene DN10 protein
33	75.5	7.8	198	2	AG0161	probable permease
34	75.5	7.8	203	2	H69903	conserved hypothet
35	75.5	7.8	204	2	A89919	conserved hypothet
36	75.5	7.8	428	2	T48284	hypothetical prote
37	75.5	7.8	437	2	H83143	hypothetical prote
38	75.5	7.8	622	2	E81793	probable inner mem
39	75	7.7	345	2	T16869	hypothetical prote
40	74.5	7.7	483	2	A38560	nitrate transport
41	74	7.6	216	2	F95056	PAP2 family protei
42	74	7.6	786	2	T02729	serine/threonine-s
43	73.5	7.6	231	2	AC1698	hypothetical prote
44	73.5	7.6	272	2	T29446	hypothetical prote
45	73.5	7.6	327	2	T33015	hypothetical prote

ALIGNMENTS

RESULT 1
E86284
F9L1.2 protein - Arabidopsis thaliana
C;Species: Arabidopsis thaliana (mouse-ear cress)
C;Date: 02-Mar-2001 #sequence_revision 02-Mar-2001 #text_change 02-Nov-2001
C;Accession: E86284
R;Theologis, A.; Ecker, J.R.; Palm, C.J.; Federspiel, N.A.; Kaul, S.; White, O.; Alonso, Chin, C.W.; Chung, M.K.; Conn, L.; Conway, A.B.; Conway, A.R.; Creasy, T.H.; Dewar, K.; ansen, N.F.; Hughes, B.; Huizar, L.
Nature 408, 816-820, 2000
A;Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.; Kim, C.
C.A.; Li, J.H.; Li, Y.; Lin, X.; Liu, S.X.; Liu, Z.A.; Luros, J.S.; Maiti, R.; Marziali, Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.
A;Authors: Salzberg, S.L.; Schwartz, J.R.; Shinn, P.; Southwick, A.M.; Sun, H.; Tallon, I ker, M.; Wu, D.; Yu, G.; Fraser, C.M.; Venter, J.C.; Davis, R.W.
A;Title: Sequence and analysis of chromosome 1 of the plant Arabidopsis.
A;Reference number: A86141; MUID:21016719; PMID:11130712
A;Accession: E86284
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-290 <STO>
A;Cross-references: GB:AE005172; NID:g5103807; PIDN:AAD39637.1; GSPDB:GN00141
C;Genetics:
A;Map position: 1

Query Match 38.4%; Score 372; DB 2; Length 290;
Best Local Similarity 47.8%; Pred. No. 7.2e-31;
Matches 75; Conservative 28; Mismatches 46; Indels 8; Gaps 4;

```
Qy         21 VIAFLSPLSLIFLAKFLKKADTRDSRQACLAASLALALNGVFTNTIKLIVGRPRPDFFYR 80
              :|| : | ::| :  |::  |  |   | |    :: : || |: ||  |||||||||:|
Db         73 LIAVVLPFAVICVYYFIRN-DVYDLHHAILGLLFSVLITGVITDAIKDAVGRPRPDFFWR 131

Qy         81 CFPDGLA--HS---DLMCTGDKDVVNEGRKSFPSGHSSFAFAGLAFASFYLAGKLHCFTP 135
              |||||:   |:   :::|||  |||| || |||||||:|::|||| | | ||:||:  |
Db        132 CFPDGIGIFHNVTKNVLCTGAKDVVKEGHKSFPSGHTSWSFAGLGFLSLYLSGKIRVF-- 189

Qy        136 QGRGKSWRFCAFLSPLLFAAVIALSRTCDYKHHWQDL 172
                ||   : |  : ||| ||:: :||  || |||||:
Db        190 DQRGHVAKLCIVILPLLVAALVGVSRVDDYWHHWQDV 226
```

RESULT 2
E84421
probable phosphatidic acid phosphatase [imported] - Arabidopsis thaliana
C;Species: Arabidopsis thaliana (mouse-ear cress)
C;Date: 02-Feb-2001 #sequence_revision 02-Feb-2001 #text_change 02-Feb-2001
C;Accession: E84421
R;Lin, X.; Kaul, S.; Rounsley, S.D.; Shea, T.P.; Benito, M.I.; Town, C.D.; Fujii, C.Y.; M.; Koo, H.; Moffat, K.S.; Cronin, L.A.; Shen, M.; VanAken, S.E.; Umayam, L.; Tallon, L. euss, D.; Nierman, W.C.; White, O.; Eisen, J.A.; Salzberg, S.L.; Fraser, C.M.; Venter, J.
Nature 402, 761-768, 1999

```
A;Title: Sequence and analysis of chromosome 2 of the plant Arabidopsis thaliana.
A;Reference number: A84420; MUID:20083487; PMID:10617197
A;Accession: E84421
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-302 <STO>
A;Cross-references: GB:AE002093; NID:g4262225; PIDN:AAD14518.1; GSPDB:GN00139
C;Genetics:
A;Gene: At2g01180
A;Map position: 2

  Query Match             34.8%;  Score 337;  DB 2;  Length 302;
  Best Local Similarity   42.1%;  Pred. No. 3.2e-27;
  Matches   69;  Conservative   31;  Mismatches   54;  Indels   10;  Gaps    4;

Qy         16 TKPMF---VIAFLSPLSLIFLAKFLKKADTRDSRQACLAASLALALNGVFTNTIKLIVGR 72
              | |::   | | | |: ::|:  :||:    |   : |    |: : || |::||:  ||
Db         65 TVPIWSVPVYAVLLPI-IVFVCFYLKRTCVYDLHHSILGLLFAVLITGVITDSIKVATGR 123

Qy         73 PRPDFFYRCFPDGL----AHSDLMCTGDKDVVNEGRKSFPSGHSSFAFAGLAFASFYLAG 128
              |||:|::||||||     |   ::| |    | || |||||||:|::|||| | | ||:|
Db        124 PRPNFYWRCFPDGKELYDALGGVVCHGKAAEVKEGHKSFPSGHTSWSFAGLTFLSLYLSG 183

Qy        129 KLHCFTPQGRGKSWRFCAFLSPLLFAAVIALSRTCDYKHHWQDL 172
              |:  |  :|     : |  : ||| | :: :||  || |||||:
Db        184 KIKAFNNEGHVA--KLCLVIFPLLAACLVGISRVDDYWHHWQDV 225


RESULT 3
S70114
probable membrane protein YDR284c - yeast (Saccharomyces cerevisiae)
N;Alternate names: hypothetical protein D9819.10
C;Species: Saccharomyces cerevisiae
C;Date: 24-Aug-1996 #sequence_revision 06-Sep-1996 #text_change 06-Feb-1998
C;Accession: S70114
R;Fulton, L.
submitted to the EMBL Data Library, May 1996
A;Description: The sequence of S. cerevisiae cosmid 9819.
A;Reference number: S70114
A;Accession: S70114
A;Molecule type: DNA
A;Residues: 1-289 <FUL>
A;Cross-references: EMBL:U51031; NID:g1332635; PID:g1332640; MIPS:YDR284c
C;Genetics:
A;Gene: SGD:DPP1
A;Cross-references: SGD:S0002692; MIPS:YDR284c
A;Map position: 4R
C;Keywords: transmembrane protein
F;19-35/Domain: transmembrane #status predicted <TM1>
F;66-82/Domain: transmembrane #status predicted <TM2>
F;92-108/Domain: transmembrane #status predicted <TM3>
F;200-216/Domain: transmembrane #status predicted <TM4>

  Query Match             25.5%;  Score 247;  DB 2;  Length 289;
  Best Local Similarity   37.8%;  Pred. No. 6.6e-18;
  Matches   68;  Conservative   27;  Mismatches   69;  Indels   16;  Gaps    6;

Qy          6 NPYVEAEYFPTKPMFVIAFLSP-LSLIFLAKFLKKADTRD----SRQACLAASLALALNG 60
              :||   |      :|| :|: | |::: :   |  || |          : |  |||
Db         54 HPYATTERVNNNMLFVYSFVVPSLTILIIGSIL--ADRRHLIFILYTSLLGLSLAWFSTS 111

Qy         61 VFTNTIKLIVGRPRPDFFYRCFP-DGL------AHSDLMCTGDKDVVNEGRKSFPSGHSS 113
               ||| ||  :|| |||| || | :||          |:  | : : : :| :: ||||||
Db        112 FFTNFIKNWIGRLRPDFLDRCQPVEGLPLDTLFTAKDVCTTKNHERLLDGFRTTPSGHSS 171

Qy        114 FAFAGLAFASFYLAGKLHCFTPQGRGKSWRFCAFLSPLLFAAVIALSRTCDYKHHWQDLL 173
              :|||| :  |:| |:|   |        ||      ||| ||:|||||| ||:||: |::
Db        172 ESFAGLGYLYFWLCGQL--LTESPLMPLWRKMVAFLPLLGAALIALSRTQDYRHHFVDVI 229


RESULT 4
T40445
phosphatidic acid phosphatase homolog - fission yeast (Schizosaccharomyces pombe)
C;Species: Schizosaccharomyces pombe
C;Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 07-Dec-1999
C;Accession: T40445
R;Lyne, M.H.; Rajandream, M.A.; Barrell, B.G.; Chillingworth, T.; Churcher, C.M.
submitted to the EMBL Data Library, August 1999
A;Reference number: Z21929
A;Accession: T40445
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-279 <LYN>
A;Cross-references: EMBL:AL109822; PIDN:CAB52620.1; GSPDB:GN00067; SPDB:SPBC409.18
A;Experimental source: strain 972h-; cosmid c409
C;Genetics:
A;Gene: SPDB:SPBC409.18
A;Map position: 2
A;Introns: 65/3

  Query Match             25.0%;  Score 242;  DB 2;  Length 279;
  Best Local Similarity   34.0%;  Pred. No. 2.1e-17;
  Matches   64;  Conservative   30;  Mismatches   56;  Indels   38;  Gaps    7;

Qy          6 NPYVEAEYFPTKPMFVIAFLSPL-----------SLIFLAKFLKKADTRDSRQACLAAS 53
              :|:   |  ||| : :|    |            ||:|              :: :
Db         51 HPFALHEQVPTKYLGIICVFFPALVLYGFGKLRNNSLLFW-------------KSLMGLL 97

Qy         54 LALALNGVFTNTIKLIVGRPRPDFFYRCFP------DGLAHSDLMCT---GDKDVVNEGR 104
               :  : |:  : :|  |||||||  || |       ||    | |:    || |: :|
Db         98 YSTMVCGLCVSLLKNAVGRPRPDFLARCQPFESTPKTGLV-DVLSCSVPWSDK-VLQDGF 155

Qy        105 KSFPSGHSSFAFAGLAFASFYLAGKLHCFTPQGRGKSWRFCAFLSPLLFAAVIALSRTCD 164
              :||||||:||:|||| | : :|||:|  |  : :  ||:    | ||  |: | |||: |
Db        156 RSFPSGHTSFSFAGLGFLAIFLAGQLKMF--RNKTSSWKVVVPLVPLSIASWIGLSRSQD 213

Qy        165 YKHHWQDL 172
              |:|| :|:
Db        214 YRHHKEDI 221


RESULT 5
S69561
hypothetical protein YDR503c - yeast (Saccharomyces cerevisiae)
C;Species: Saccharomyces cerevisiae
C;Date: 22-Aug-1996 #sequence_revision 06-Sep-1996 #text_change 19-Apr-2002
C;Accession: S69561
R;Dietrich, F.S.
submitted to the EMBL Data Library, August 1995
A;Description: The sequence of S. cerevisiae cosmids 8166, 9787, 9717, and lambda 3073.
A;Reference number: S69553
A;Accession: S69561
A;Molecule type: DNA
A;Residues: 1-274 <DIE>
A;Cross-references: EMBL:U33057; NID:g927764; PID:g927773; GSPDB:GN00004; MIPS:YDR503c
C;Genetics:
A;Gene: SGD:LPP1; MIPS:YDR503c
A;Cross-references: SGD:S0002911
A;Map position: 4R

  Query Match             20.7%;  Score 200.5;  DB 2;  Length 274;
  Best Local Similarity   29.0%;  Pred. No. 4.1e-13;
  Matches   60;  Conservative   33;  Mismatches   59;  Indels   55;  Gaps   10;

Qy          6 NPYVEAEYFPTKPMFVIAFLSPLSLIFLA---------------KFLKKADTRDSRQ-- 47
              :| :   | | :       : ||  : |:                  |||   :  |:
Db         51 DPSISKRYVPNE------LVGPLECLILSVGLSNMVVFWTCMFDKDLLKKNRVKRLRERP 104

Qy         48 ------------ACLAASLALALNGVFTNTIKLIVGRPRPDFFYRCFPD--GLAHSDLMC 93
                         : |  | :::|   |  :|||:|  ||||  || ||   :: || :
Db        105 DGISNDFHFMHTSILCLMLIISINAALTGALKLIIGNLRPDFVDRCIPDLQKMSDSDSLV 164

Qy         94 TG-------DKDVVNEGRKSFPSGHSSFAFAGLAFASFYLAGKLHCFTPQGRGKSWRFCA 146
                |       :| :: || || |||||||  : : |   ||  ::  ||    :: | |
Db        165 FGLDICKQTNKWILYEGLKSTPSGHSSFIVSTMGFT--YLWQRV--FTT----RNTRSCI 216
```

```
Qy        147 FLSPLLFAAVIALSRTCDYKHHWQDLL 173
              :  ||| | |: :||  |::||| |::
Db        217 W-CPLL-ALVVMVSRVIDHRHHWYDVV 241
```

RESULT 6
T16951
hypothetical protein T28D9.3 - Caenorhabditis elegans
C;Species: Caenorhabditis elegans
C;Date: 20-Sep-1999 #sequence_revision 20-Sep-1999 #text_change 20-Sep-1999
C;Accession: T16951
R;Fulton, L.
submitted to the EMBL Data Library, June 1995
A;Description: The sequence of C. elegans cosmid T28D9.
A;Reference number: Z18614
A;Accession: T16951
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-341 <FUL>
A;Cross-references: EMBL:U28738; NID:g861262; PID:g861267; PIDN:AAA68312.1; CESP:T28D9.3
A;Experimental source: strain Bristol N2
C;Genetics:
A;Gene: CESP:T28D9.3
A;Introns: 87/3; 127/1; 172/2; 209/3; 269/3; 303/1

  Query Match 14.8%; Score 143; DB 2; Length 341;
  Best Local Similarity 30.6%; Pred. No. 4.8e-07;
  Matches 41; Conservative 20; Mismatches 55; Indels 18; Gaps 4;

```
Qy         52 ASLALALNGVFTNTIKLIVGRPRPDFFYRC------FPDGLAH---SDLMCTGDKDVVNE 102
              : :  :|       | :||| || |   |         | :|   :|  |||  ::| |
Db        129 SQIGFVMNIALNIVTKHVVGRLRPHFLDVCKLANDTCVTGDSHRYITDYTCTGPPELVLE 188

Qy        103 GRKSFPSGHSSFAFAGLAFASFYLAGKLHCFTPQGRGKSWRFCAFLSPLLFAAV---IAL 159
               |||| ||||: :     ::: |:  :|      |   : |    :|  |   :   |:
Db        189 ARKSFYSGHSAVSLYCATWSALYIQARL------GPVLNNRIVVPISQTLMFMIGLGISF 242

Qy        160 SRTCDYKHHWQDLL 173
              ||  | |||| |:|
Db        243 SRITDNKHHWSDVL 256
```

RESULT 7
S64327
probable membrane protein YGR036c - yeast (Saccharomyces cerevisiae)
N;Alternate names: hypothetical protein G4085
C;Species: Saccharomyces cerevisiae
C;Date: 17-May-1996 #sequence_revision 17-May-1996 #text_change 06-Feb-1998
C;Accession: S64327
R;Rieger, M.; Mueller-Auer, S.; Brueckner, M.; Schaefer, M.
submitted to the Protein Sequence Database, May 1996
A;Reference number: S64071
A;Accession: S64327
A;Molecule type: DNA
A;Residues: 1-239 <RIE>
A;Cross-references: EMBL:Z72821; NID:g1323018; PID:e243385; PID:g1323019; MIPS:YGR036c
A;Experimental source: strain S288C
C;Genetics:
A;Gene: SGD:CWH8
A;Cross-references: SGD:S0003268; MIPS:YGR036c
A;Map position: 7R
C;Keywords: transmembrane protein
F;39-55/Domain: transmembrane #status predicted <TM1>
F;136-152/Domain: transmembrane #status predicted <TM2>
F;169-185/Domain: transmembrane #status predicted <TM3>

  Query Match 12.8%; Score 124; DB 2; Length 239;
  Best Local Similarity 27.8%; Pred. No. 3.1e-05;
  Matches 42; Conservative 21; Mismatches 50; Indels 38; Gaps 8;

```
Qy         13 YFPTKPMFVIAFLSPLSLIFLAKFLKKADTRDSRQACLAASLALALNGVFTNTIKLIVGR 72
              ||   |: |:||       :|: |:    ||:  :||: |    :| :| | || |: :
Db         36 YFSLMPILVLAF-------YLSWFI---ITRE-LEACIVA-FGQLMNEIFNNVIKNIIKQ 83

Qy         73 PRPDFFYRCFPDGLAHSDLMCTGDKDVVNEGRKSFPSGHSSFAFAGLAFASFYLAGKLHC 132
              |||  |   |               | :  |    || || |    : |   | : |::
Db         84 PRPVSFGASF------------QNDTIRSG-YGMPSAHSQF----MGFCFTYNSLKIY- 124

Qy        133 FTPQGRGKSWRFCAFLSPLLFAAVIALSRTC 163
                 ||:   ||   :|:  :||   |
Db        125 -------TSWKNLNFLEKCIFSGALALLSFC 148
```

RESULT 8
JE0199
phosphatidic acid phosphatase (EC 3.1.-.-) - Caenorhabditis elegans
N;Alternate names: PAP
C;Species: Caenorhabditis elegans
C;Date: 21-Aug-1998 #sequence_revision 21-Aug-1998 #text_change 07-May-1999
C;Accession: JE0199
R;Katagiri, T.; Shinozaki, K.
Biochem. Biophys. Res. Commun. 248, 87-92, 1998
A;Title: Disruption of a gene encoding phosphatidic acid phosphatase causes abnormal pher
A;Reference number: JE0199; MUID:98340855; PMID:9675091
A;Accession: JE0199
A;Molecule type: mRNA
A;Residues: 1-318 <KAT>
A;Cross-references: EMBL:Z68105
C;Comment: This enzyme is involved in lipid metabolism.
C;Keywords: hydrolase

  Query Match 12.2%; Score 118; DB 2; Length 318;
  Best Local Similarity 28.5%; Pred. No. 0.00017;
  Matches 39; Conservative 14; Mismatches 52; Indels 32; Gaps 4;

```
Qy         17 KPMFVIAFLSPLSLIF--------------LAKFLKKADTRDSRQACLAASLALALNGV 61
              | : ||   ||  ::               ||||         :   : |:   |:
Db        101 KHLLVITLGSPFLIVALVEAILHFKSKGSNRLAKFFSATTITYLKYLLMYAACTFAM--- 157

Qy         62 FTNTIKLIVGRPRPDFFYRCFPDG----------LAHSDLMCTG-DKDVVNEGRKSFPS 109
                 :|  ||| || ||  | ||           :  |||:||  :   :   | ||||
Db        158 --EFLKCYVGRLRPHFFSVCKPDWSKVDCTDKQSFIDSSDLVCTNPNPRKIRTARTSFPS 215

Qy        110 GHSSFAFAGLAFASFYL 126
              ||:: ||    |   ||
Db        216 GHTAAAFHVFLFVYIYL 232
```

RESULT 9
H89632
protein F13E6.5 [imported] - Caenorhabditis elegans
C;Species: Caenorhabditis elegans
C;Date: 10-May-2001 #sequence_revision 10-May-2001 #text_change 09-Nov-2001
C;Accession: H89632
R;anonymous, The C. elegans Sequencing Consortium.
Science 282, 2012-2018, 1998
A;Title: Genome sequence of the nematode C. elegans: a platform for investigating biology
A;Reference number: A75000; MUID:99069613; PMID:9851916
A;Note: see websites genome.wustl.edu/gsc/C_elegans/ and www_sanger.ac.uk/Projects/C_eleg
A;Note: published errata appeared in Science 283, 35, 1999; Science 283, 2103, 1999; and
A;Accession: H89632
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-318 <STO>
A;Cross-references: GB:chr_X; PIDN:CAA92120.1; PID:g3875840; GSPDB:GN00028; CESP:F13E6.5
C;Genetics:
A;Gene: F13E6.5
A;Map position: X

  Query Match 12.2%; Score 118; DB 2; Length 318;
  Best Local Similarity 28.5%; Pred. No. 0.00017;
  Matches 39; Conservative 14; Mismatches 52; Indels 32; Gaps 4;

```
Qy         17 KPMFVIAFLSPLSLIF--------------LAKFLKKADTRDSRQACLAASLALALNGV 61
              | : ||   ||  ::               ||||         :   : |:   |:
Db        101 KHLLVITLGSPFLIVALVEAILHFKSKGSNRLAKFFSATTITYLKYLLMYAACTFAM--- 157
```

```
Qy          62 FTNTIKLIVGRPRPDFFYRCFPDG-----------LAHSDLMCTG-DKDVVNEGRKSFPS 109
Db         158 --EFLKCYVGRLRPHFFSVCKPDWSKVDCTDKQSFIDSSDLVCTNPNPRKIRTARTSFPS 215

Qy         110 GHSSFAFAGLAFASFYL 126
Db         216 GHTAAAFHVFLFVYIYL 232
```

RESULT 10
D69206
conserved hypothetical protein MTH798 - Methanobacterium thermoautotrophicum (strain Del
C;Species: Methanobacterium thermoautotrophicum
C;Date: 28-Jan-2000 #sequence_revision 28-Jan-2000 #text_change 03-Mar-2000
C;Accession: D69206
R;Smith, D.R.; Doucette-Stamm, L.A.; Deloughery, C.; Lee, H.; Dubois, J.; Aldredge, T.;
; Qiu, D.; Spadafora, R.; Vicaire, R.; Wang, Y.; Wierzbowski, J.; Gibson, R.; Jiwani, N.
ki, S.; Church, G.M.; Daniels, C.J.; Mao, J.; Rice, P.; Noelling, J.; Reeve, J.N.
J. Bacteriol. 179, 7135-7155, 1997
A;Title: Complete genome sequence of Methanobacterium thermoautotrophicum Delta H: funct
A;Reference number: A69000; MUID:98037514; PMID:9371463
A;Accession: D69206
A;Status: preliminary; nucleic acid sequence not shown; translation not shown
A;Molecule type: DNA
A;Residues: 1-186 <MTH>
A;Cross-references: GB:AE000858; GB:AE000666; NID:g2621885; PIDN:AAB85298.1; PID:g262188
A;Experimental source: strain Delta H
C;Genetics:
A;Gene: MTH798
C;Superfamily: bacitracin transport permease; glucose-6-phosphatase catalytic domain hom
C;Keywords: antibiotic resistance; transmembrane protein
F;55-161/Domain: glucose-6-phosphatase catalytic domain homology <GPH>

  Query Match   11.9%; Score 115.5; DB 1; Length 186;
  Best Local Similarity   26.3%; Pred. No. 0.00018;
  Matches   45; Conservative   20; Mismatches   59; Indels   47; Gaps   8;

```
Qy          24 FLSPLSL--------IFLAKFLKKADTRDSRQACLAASLALALNGVFTNTIKLIVGRPRP 75
Db          27 FLMPLLTFGGTQAFWVILCLLLYLLGGEDEREAAFMALTALVLGFFLSEALKMVIARPRP 86

Qy          76 DFFYRCFPDGLAHSDLMCTGDKDVVNEGRKSFPSGHSSFAFAGLAFASFYLAGKLHCFTP 135
Db          87 ---YEVI-GWVRHATV----------AGGYSMPSGHAVAAFAG--FISLYFR-------- 122

Qy         136 QGRGKSWRFCAFLSPLLFAAVIALSRTCDYKHHWQDLLK-------CTNTA 179
Db         123 --LGRPWFF------IILASLVGISRIYLGLHYPSDVLAGAVLGVLCAFTA 165
```

RESULT 11
E86750
hypothetical protein ykdB [imported] - Lactococcus lactis subsp. lactis (strain IL1403)
C;Species: Lactococcus lactis subsp. lactis
C;Date: 23-Mar-2001 #sequence_revision 23-Mar-2001 #text_change 12-May-2003
C;Accession: E86750
R;Bolotin, A.; Wincker, P.; Mauger, S.; Jaillon, O.; Malarme, K.; Weissenbach, J.; Ehrli
Genome Res. 11, 731-753, 2001
A;Title: The complete genome sequence of the lactic acid bacterium Lactococcus lactis ss
A;Reference number: A86625; MUID:21235186; PMID:11337471
A;Accession: E86750
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-218 <STO>
A;Cross-references: GB:AE005176; PID:g12723952; PIDN:AAK05103.1; GSPDB:GN00146
A;Experimental source: strain IL1403
C;Genetics:
A;Gene: ykdB
C;Superfamily: bacitracin transport permease

  Query Match   11.6%; Score 112; DB 2; Length 218;
  Best Local Similarity   30.2%; Pred. No. 0.00049;
  Matches   49; Conservative   18; Mismatches   59; Indels   36; Gaps   7;

```
Qy          14 FPTKPMFVIAFLSPLSLIFLAKFLKKADTRDSRQACLAAS--LALALNGVFTNTIKLIVG 71
Db          63 FGDKGGIIVAVL--LALVLYLLFKQKEAT-----IWFAATVVLSLALNTV----IKAIIG 111

Qy          72 RPRPDFFYRCFPDGLAHSDLMCTGDKDVVNEGRKSFPSGHSSFAFAGLAFASFYLAGKLH 131
Db         112 RERPD-IHRL---------------AAFANEAGKSFPSGHSIFATIIFGSIFFICLGKL- 154

Qy         132 CFTPQGRGKSWRFCAFLSPLLFAAVIALSRTCDYKHHWQDLL 173
Db         155 ------KNRSSKILLGILCIILIALIMFSRIFVGVHYPSDTI 190
```

RESULT 12
D97672
hypothetical protein AGR_C_4732 [imported] - Agrobacterium tumefaciens (strain C58, Cerec
C;Species: Agrobacterium tumefaciens
C;Date: 30-Sep-2001 #sequence_revision 30-Sep-2001 #text_change 12-May-2003
C;Accession: D97672
R;Goodner, B.; Hinkle, G.; Gattung, S.; Miller, N.; Blanchard, M.; Qurollo, B.; Goldman,
 A.; Liu, F.; Wollam, C.; Allinger, M.; Doughty, D.; Scott, C.; Lappas, C.; Markelz, B.;
Science 294, 2323-2328, 2001
A;Title: Genome Sequence of the Plant Pathogen and Biotechnology Agent Agrobacterium tume
A;Reference number: A97359; MUID:21608551; PMID:11743194
A;Accession: D97672
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-217 <KUR>
A;Cross-references: GB:AE007869; PIDN:AAK88333.1; PID:g15157811; GSPDB:GN00169
C;Genetics:
A;Gene: AGR_C_4732
A;Map position: circular chromosome
C;Superfamily: bacitracin transport permease

  Query Match   10.7%; Score 104; DB 2; Length 217;
  Best Local Similarity   31.1%; Pred. No. 0.0033;
  Matches   37; Conservative   15; Mismatches   41; Indels   26; Gaps   6;

```
Qy          54 LALALNGVFTNTIKLIVGRPRPDFFYRCFPDGLAHSDLMCTGDKDVVNEGRKSFPSGHSS 113
Db          79 LTIGTASILVHTLKFLIGRARPELFLEMG----AYSLTPFTGDNLY-----ESFPSGHST 129

Qy         114 FAFAGLAFASFYLAGKLHCFTPQGRGKSWRFCAFLSPLLFAAVIALSRTCDYKHHWQDL 172
Db         130 --AAGAFFGVFAM------LMPRFR---WAF------LLLALVIGVSRVIVGAHYPSDV 171
```

RESULT 13
AB2897
conserved hypothetical protein Atu2611 [imported] - Agrobacterium tumefaciens (strain C58
C;Species: Agrobacterium tumefaciens
C;Date: 11-Jan-2002 #sequence_revision 11-Jan-2002 #text_change 12-May-2003
C;Accession: AB2897
R;Wood, D.W.; Setubal, J.C.; Kaul, R.; Monks, D.; Chen, L.; Wood, G.E.; Chen, Y.; Woo, L.
erage, G.; Gillet, W.; Grant, C.; Guenthner, D.; Kutyavin, T.; Levy, R.; Li, M.; McClella
; Karp, P.; Romero, P.; Zhang, S.
Science 294, 2317-2323, 2001
A;Authors: Yoo, H.; Tao, Y.; Biddle, P.; Jung, M.; Krespan, W.; Perry, M.; Gordon-Kamm, B
ster, E.W.
A;Title: The Genome of the Natural Genetic Engineer Agrobacterium tumefaciens C58.
A;Reference number: AB2577; MUID:21608550; PMID:11743193
A;Accession: AB2897
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-242 <KUR>
A;Cross-references: GB:AE008688; PIDN:AAL43592.1; PID:g17741109; GSPDB:GN00186
A;Experimental source: strain C58 (Dupont)
C;Genetics:
A;Gene: Atu2611
A;Map position: circular chromosome
C;Superfamily: bacitracin transport permease

```
  Query Match             10.7%;  Score 104;  DB 2;  Length 242;
  Best Local Similarity   31.1%;  Pred. No. 0.0037;
  Matches   37;  Conservative   15;  Mismatches   41;  Indels   26;  Gaps    6;

Qy         54 LALALNGVFTNTIKLIVGRPRPDFFYRCFPDGLAHSDLMCTGDKDVVNEGRKSFPSGHSS 113
              | :    :  :|:| ::|| ||: |        |:|    |||        :|||||||:
Db        104 LTIGTASILVHTLKFLIGRARPELFLEMG----AYSLTPFTGDNLY-----ESFPSGHST 154

Qy        114 FAFAGLAFASFYLAGKLHCFTPQGRGKSWRFCAFLSPLLFAAVIALSRTCDYKHHWQDL 172
              ||  |  | :       |: |   | |      || | || :||     |:  |:
Db        155 --AAGAFFGVFAM------LMPRFR---WAF------LLLALVIGVSRVIVGAHYPSDV 196


RESULT 14
T24576
hypothetical protein T06D8.3 - Caenorhabditis elegans
C;Species: Caenorhabditis elegans
C;Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 15-Oct-1999
C;Accession: T24576
R;Palmer, S.
submitted to the EMBL Data Library, April 1995
A;Reference number: Z19909
A;Accession: T24576
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-396 <WIL>
A;Cross-references: EMBL:Z49130; PIDN:CAA88966.1; GSPDB:GN00020; CESP:T06D8.3
A;Experimental source: clone T06D8
C;Genetics:
A;Gene: CESP:T06D8.3
A;Map position: 2
A;Introns: 11/3; 59/2; 184/2; 205/2; 233/3; 292/3; 340/2

  Query Match             10.6%;  Score 102.5;  DB 2;  Length 396;
  Best Local Similarity   23.7%;  Pred. No. 0.0089;
  Matches   50;  Conservative   29;  Mismatches   81;  Indels   51;  Gaps    9;

Qy          3 LYRNPYVEAEY--FPTKP-MFVIAFLSPLSLIFLAK-----FLKKAD------------ 41
              ||:  :|  ::  : : | :: :||  |  :| : :     |  |
Db         78 LYKPNFVPEDFNVYVSYPLLYTLAFTIPPLVILIGEVMFWLFSTKPRKIVYANCGECPVH 137

Qy         42 --TRDSRQACLAASLALALNGVFTNTIKLIVGRPRPDFFYRC------FPDGLAHS---- 89
                ||   :  :      | :  :| :||||: |  || |   |           | ||
Db        138 LFTRRLFRFVIIYLAGLLIVQIFVDTIKLMTGYQRPYFLSLCNVSITACTAPLEHSPSPS 197

Qy         90 -DLMCT-GDKDVVNEGRKSFPSGHSSFAFAGLAFASFYLAGKLHCFTPQGRGKSWRFCAF 147
               | |     | :     :||| |:  :     ||| |:   ::           |
Db        198 PHLACNYRGADELRYAWLTFPSLHAVVSSYAACFASLYIYYMINL----------RGAPL 247

Qy        148 LSPLLFAAVIAL------SRTCDYKHHWQDL 172
              | |||    : |      ||   ||:||:|:
Db        248 LRPLLIFGFMGLCIVDSFSRINGYKNHWRDI 278


RESULT 15
AF3647
phosphatidylglycerophosphatase (EC 3.1.3.27) [imported] - Brucella melitensis (strain 16
C;Species: Brucella melitensis
C;Date: 01-Feb-2002 #sequence_revision 01-Feb-2002 #text_change 03-Jun-2002
C;Accession: AF3647
R;DelVecchio, V.G.; Kapatral, V.; Redkar, R.J.; Patra, G.; Mujer, C.; Los, T.; Ivanova,
.; Mazur, M.; Goltsman, E.; Selkov, E.; Elzer, P.H.; Hagius, S.; O'Callaghan, D.; Letess
Proc. Natl. Acad. Sci. U.S.A. 99, 443-448, 2002
A;Title: The genome sequence of the facultative intracellular pathogen Brucella melitens
A;Reference number: AD3252; PMID:11756688
A;Accession: AF3647
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-292 <KUR>
A;Cross-references: GB:AE008918; PIDN:AAL54345.1; PID:g17985327; GSPDB:GN00191
A;Experimental source: strain 16M
C;Genetics:
A;Gene: BMEII1103
A;Map position: II
C;Keywords: phosphoric monoester hydrolase

  Query Match             10.1%;  Score 97.5;  DB 2;  Length 292;
  Best Local Similarity   26.1%;  Pred. No. 0.021;
  Matches   41;  Conservative   25;  Mismatches   50;  Indels   41;  Gaps   10;

Qy         24 FLSPLSLIFL-AKFLKKADTRDSRQACLAASLALALNGVFTNTIKLIVGRPRP-----DF 77
              || | : :|:  :|: ::  :      |      ::|::|:    :|:| || ||     |
Db         37 FLVPAAALFVFYRFINRSPQKSFN--CFFIIASIAVSGIVIKILKIIFGRARPGVLIDDG 94

Qy         78 FYRCFPDGLAHSDLMCTGDKDVVNEGRKSFPSGHSSFAF-AGLAFASFYLAGKLHCFTPQ 136
              ||     |     |    |::       |||| |:  |  ||:| |   |  : |
Db         95 FY-----GFTFFRL----DREF-----NSFPSAHTGVAIAAGVALA---LIMQKH----- 132

Qy        137 GRGKSWRFCAFLSPLLFAAVIALSRTCDYKHHWQDLL 173
                     |       |::   ||| ||      |:  |::
Db        133 ----RW------VPIILGIVIASSRIIINAHYLSDVV 159


Search completed: August 14, 2003, 13:08:46
Job time : 41 secs
```

GenCore version 5.1.6

OM protein - protein search, using sw model

Run on: August 14, 2003, 13:03:36 ; Search time 22 Seconds
(without alignments)
384.764 Million cell updates/sec

Title: US-09-876-997-399
Perfect score: 968
Sequence: 1 MWLYRNPYVEAEYFPTKPMF.....RTCDYKHHWQDLLKCTNTAK 180

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 127863 seqs, 47026705 residues

Total number of hits satisfying chosen parameters: 127863

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SwissProt_41:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	172	17.8	379	1	WUN_DROME	Q9v576 drosophila
2	143	14.8	341	1	YSX3_CAEEL	Q10022 caenorhabdi
3	124	12.8	239	1	YG1P_YEAST	P53223 saccharomyc
4	90.5	9.3	330	1	Y374_METJA	Q57819 methanococc
5	77.5	8.0	378	1	YF69_AQUAE	O67513 aquifex aeo
6	76.5	7.9	838	1	P_HUMAN	Q04671 homo sapien
7	74.5	7.7	507	1	CRNA_EMENI	P22152 emericella
8	73	7.5	921	1	ITH4_PIG	P79263 sus scrofa
9	70.5	7.3	576	1	ML12_ARATH	O80961 arabidopsis
10	70	7.2	133	1	E1BS_ADECT	P14265 canine aden
11	70	7.2	507	1	UDP2_PIG	P79303 sus scrofa
12	69	7.1	1704	1	ABC3_HUMAN	Q99758 homo sapien
13	69	7.1	2233	1	RRPL_PI3H4	P12577 human parai
14	68	7.0	200	1	ISPZ_BRAJA	P30961 bradyrhizob
15	68	7.0	276	1	YGFR_ECOLI	Q46818 escherichia
16	67.5	7.0	237	1	YEIU_ECOLI	P76445 escherichia
17	67.5	7.0	444	1	CRF1_HUMAN	P34998 homo sapien
18	67.5	7.0	551	1	FZD2_XENLA	Q9puu6 xenopus lae
19	67.5	7.0	1313	1	MIP1_SCHPO	P87141 schizosacch
20	67	6.9	305	1	ATNB_TORCA	P05029 torpedo cal
21	67	6.9	826	1	RIR1_EBV	P03190 epstein-bar
22	67	6.9	938	1	PM15_CHLPN	Q9z883 chlamydia p
23	66.5	6.9	153	1	VE6_HPV57	P22158 human papil
24	66.5	6.9	381	1	CYB_CHLRE	P23662 chlamydomon
25	66.5	6.9	381	1	CYB_CHLSM	P23663 chlamydomon
26	66.5	6.9	704	1	PNP_BACSU	P50849 bacillus su
27	66.5	6.9	892	1	TNPA_STRFR	P20189 streptomyce
28	66.5	6.9	2139	1	CCAC_MOUSE	Q01815 mus musculu
29	66.5	6.9	2169	1	CCAC_RAT	P22002 rattus norv
30	66.5	6.9	2221	1	CCAC_HUMAN	Q13936 homo sapien
31	66	6.8	293	1	MT16_EMENI	P56859 emericella
32	66	6.8	344	1	GCP_MYCTU	Q50709 mycobacteri
33	66	6.8	658	1	KPC1_LYTPI	Q25378 lytechinus
34	66	6.8	1379	1	YFF9_SCHPO	O14066 schizosacch
35	65.5	6.8	177	1	CCAC_CHICK	O73707 gallus gall
36	65.5	6.8	380	1	METB_HELPY	P56069 helicobacte
37	65.5	6.8	433	1	YBL4_YEAST	P38211 saccharomyc
38	65.5	6.8	1152	1	YBF7_CAEEL	Q21286 caenorhabdi
39	65.5	6.8	2171	1	CCAC_RABIT	P15381 oryctolagus
40	64.5	6.7	247	1	LPXH_XANCP	P58976 xanthomonas
41	64.5	6.7	326	1	TRXB_BORBU	P94284 borrelia bu
42	64.5	6.7	498	1	YM22_CAEEL	Q21339 caenorhabdi
43	64	6.6	255	1	THID_AQUAE	O67772 aquifex aeo
44	64	6.6	368	1	PTIC_ECOLI	P31452 escherichia
45	64	6.6	415	1	CRF1_MOUSE	P35347 mus musculu

ALIGNMENTS

RESULT 1
WUN_DROME
ID WUN_DROME STANDARD; PRT; 379 AA.
AC Q9V576; P91661; Q8MKU5; Q8WT60; Q9U9Y7;
DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 15-SEP-2003 (Rel. 42, Last annotation update)
DE Putative phosphatidate phosphatase (EC 3.1.3.4) (Phosphatidic acid
DE phosphatase type 2) (Wunen protein) (Germ cell guidance factor).
GN WUN OR CG8804.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN [1]
RP SEQUENCE FROM N.A. (ISOFORM SHORT), FUNCTION, AND TISSUE SPECIFICITY.
RX MEDLINE=97138220; PubMed=8985246;
RA Zhang N., Zhang J.P., Purcell K.J., Chen Y., Howard K.;
RT "The Drosophila protein Wunen repels migrating germ cells.";
RL Nature 385:64-67(1997).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=Berkeley;
RX MEDLINE=20196006; PubMed=10731132;
RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA Brandon R.C., Rogers Y.-H.C., Blazej R.G., Champe M., Pfeiffer B.D.,
RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
RA Abril J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brottier P.,
RA Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA Durbin K.J., Evangelista C.C., Ferraz C., Ferriera S., Fleischmann W.,
RA Fosler C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,
RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleb J.M.,
RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,

```
RA   Wang Z.-Y., Wassarman D.A., Weinstock G.M., Weissenbach J.,
RA   Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
RA   Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA   Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA   Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RT   "The genome sequence of Drosophila melanogaster.";
RL   Science 287:2185-2195(2000).
RN   [3]
RP   REVISIONS, AND ALTERNATIVE SPLICING.
RC   STRAIN=Berkeley;
RX   MEDLINE=22426069; PubMed=12537572;
RA   Misra S., Crosby M.A., Mungall C.J., Matthews B.B., Campbell K.S.,
RA   Hradecky P., Huang Y., Kaminker J.S., Millburn G.H., Prochnik S.E.,
RA   Smith C.D., Tupy J.L., Whitfield E.J., Bayraktaroglu L., Berman B.P.,
RA   Bettencourt B.R., Celniker S.E., de Grey A.D.N.J., Drysdale R.A.,
RA   Harris N.L., Richter J., Russo S., Schroeder A.J., Shu S.Q.,
RA   Stapleton M., Yamada C., Ashburner M., Gelbart W.M., Rubin G.M.,
RA   Lewis S.E.;
RT   "Annotation of the Drosophila melanogaster euchromatic genome: a
RT   systematic review.";
RL   Genome Biol. 3:RESEARCH0083.1-RESEARCH0083.22(2002).
RN   [4]
RP   SEQUENCE FROM N.A. (ISOFORM SHORT).
RC   STRAIN=Berkeley; TISSUE=Head;
RX   MEDLINE=20196012; PubMed=10731138;
RA   Rubin G.M., Hong L., Brokstein P., Evans-Holm M., Frise E.,
RA   Stapleton M., Harvey D.A.;
RT   "A Drosophila complementary DNA resource.";
RL   Science 287:2222-2224(2000).
RN   [5]
RP   SEQUENCE FROM N.A. (ISOFORM LONG).
RC   STRAIN=Berkeley; TISSUE=Embryo;
RX   MEDLINE=22426066; PubMed=12537569;
RA   Stapleton M., Carlson J.W., Brokstein P., Yu C., Champe M.,
RA   George R.A., Guarin H., Kronmiller B., Pacleb J.M., Park S., Wan K.H.,
RA   Rubin G.M., Celniker S.E.;
RT   "A Drosophila full-length cDNA resource.";
RL   Genome Biol. 3:RESEARCH0080.1-RESEARCH0080.8(2002).
RN   [6]
RP   SEQUENCE OF 22-99 FROM N.A.
RA   Russell C., Bartos S., Phillips R.G., Whittle R.;
RT   "Efficient functional dissection of enhancer sequences in Drosophila:
RT   a novel P- and hobo-based construct acting as an enhancer-capture
RT   element.";
RL   Submitted (JUL-2001) to the EMBL/GenBank/DDBJ databases.
CC   -!- FUNCTION: Responsible for guiding the germ cells early in the
CC       process of migration from the lumen of the developing gut towards
CC       the overlying mesoderm, where the germ cells enter the gonads. May
CC       be involved in lipid metabolism.
CC   -!- CATALYTIC ACTIVITY: A 3-sn-phosphatidate + H(2)O = a 1,2-diacyl-
CC       sn-glycerol + phosphate.
CC   -!- SUBCELLULAR LOCATION: Integral membrane protein (Potential).
CC   -!- ALTERNATIVE PRODUCTS:
CC       Event=Alternative splicing; Named isoforms=2;
CC       Name=Long;
CC         IsoId=Q9V576-1; Sequence=Displayed;
CC       Name=Short;
CC         IsoId=Q9V576-2; Sequence=VSP_005084;
CC         Note=No experimental confirmation available;
CC   -!- TISSUE SPECIFICITY: Expressed in embryonic gut in a pattern that
CC       guides germ cells towards mesoderm (initially in hindgut and then
CC       on lower side of gut). During extended germ band stage, expressed
CC       in ectoderm as a medial band throughout the trunk.
CC   -!- SIMILARITY: BELONGS TO THE PHOSPHATIDATE PHOSPHATASE FAMILY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; U73822; AAC47449.1; -.
DR   EMBL; AE003833; AAF58942.2; -.
DR   EMBL; AE003833; AAM71066.1; -.
DR   EMBL; AF145595; AAD38570.1; -.
DR   EMBL; BT001729; AAN71484.1; -.
DR   EMBL; AY046533; AAL34392.1; -.
DR   FlyBase; FBgn0016078; wun.
DR   GO; GO:0016021; C:integral to membrane; NAS.
DR   GO; GO:0008195; F:phosphatidate phosphatase activity; NAS.
DR   GO; GO:0016311; P:dephosphorylation; NAS.
DR   GO; GO:0008354; P:germ-cell migration; IEP.
DR   InterPro; IPR000326; PA_PTPase.
DR   Pfam; PF01569; PAP2; 1.
DR   SMART; SM00014; acidPPc; 1.
KW   Hydrolase; Developmental protein; Glycoprotein; Transmembrane;
KW   Alternative splicing.
FT   TRANSMEM     90    110       POTENTIAL.
FT   TRANSMEM    138    158       POTENTIAL.
FT   TRANSMEM    266    286       POTENTIAL.
FT   TRANSMEM    330    350       POTENTIAL.
FT   CARBOHYD     51     51       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    169    169       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   VARSPLIC      1     79       Missing (in isoform Short).
FT                                /FTId=VSP_005084.
FT   CONFLICT    152    152       F -> L (IN REF. 4).
FT   CONFLICT    360    360       A -> G (IN REF. 4).
SQ   SEQUENCE   379 AA;  42789 MW;  AD3C4F06B3DD9ADA CRC64;

  Query Match           17.8%;  Score 172;  DB 1;  Length 379;
  Best Local Similarity  34.1%;  Pred. No. 1.1e-10;
  Matches   47;  Conservative   14;  Mismatches   45;  Indels   32;  Gaps    5;

Qy         58 LNGVFTNTIKLIVGRPRPDFFYRCFP---DGLAHSDLM----------CTG---DKDVVN 101
              |: : |:  |  :|| || |  | |   ||   | :          | |      ::
Db        202 LSQLTTDIAKYSIGRLRPHFIAVCQPQMADGSTCDDAINAGKYIQEFTCKGVGSSARMLK 261

Qy        102 EGRKSFPSGHSSFAFAGLAFASFYLAGKLHCFTPQGRGKSWRFCAFLSPLL------FAA 155
              | | |||||||| |  : : : ||  ::          :||    |  ||      |
Db        262 EMRLSFPSGHSSFTFFAMVYLALYLQARM----------TWRGSKLLRHLLQFLFIMVAW 311

Qy        156 VIALSRTCDYKHHWQDLL 173
                |||| |||||| |:|
Db        312 YTALSRVSDYKHHWSDVL 329


RESULT 2
YSX3_CAEEL
ID   YSX3_CAEEL     STANDARD;      PRT;   341 AA.
AC   Q10022;
DT   01-OCT-1996 (Rel. 34, Created)
DT   01-OCT-1996 (Rel. 34, Last sequence update)
DT   01-OCT-1996 (Rel. 34, Last annotation update)
DE   Hypothetical 39.0 kDa protein T28D9.3 in chromosome II.
GN   T28D9.3.
OS   Caenorhabditis elegans.
OC   Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC   Rhabditidae; Peloderinae; Caenorhabditis.
OX   NCBI_TaxID=6239;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=Bristol N2;
RA   Waterston R.;
RL   Submitted (JUN-1995) to the EMBL/GenBank/DDBJ databases.
CC   -!- SUBCELLULAR LOCATION: Integral membrane protein (Potential).
CC   -!- SIMILARITY: WEAK, TO YEAST D9719.9.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
```

```
CC   ----------------------------------------------------------------------
DR   EMBL; U28738; AAA68312.1; -.
DR   PIR; T16951; T16951.
DR   WormPep; T28D9.3; CE02068.
DR   InterPro; IPR000326; PA_PTPase.
DR   Pfam; PF01569; PAP2; 1.
DR   SMART; SM00014; acidPPc; 1.
KW   Hypothetical protein; Transmembrane.
FT   TRANSMEM     30     50       POTENTIAL.
FT   TRANSMEM     71     91       POTENTIAL.
FT   TRANSMEM    122    142       POTENTIAL.
FT   TRANSMEM    223    243       POTENTIAL.
FT   TRANSMEM    257    277       POTENTIAL.
SQ   SEQUENCE   341 AA;  39028 MW;  DF74E39BC3E2DA8E CRC64;

  Query Match             14.8%;  Score 143;  DB 1;  Length 341;
  Best Local Similarity   30.6%;  Pred. No. 1.2e-07;
  Matches   41;  Conservative   20;  Mismatches   55;  Indels   18;  Gaps    4;

Qy         52 ASLALALNGVFTNTIKLIVGRPRPDFFYRC------FPDGLAH---SDLMCTGDKDVVNE 102
              : :   :|       | :||| || |   |          | :|   :|  |||  ::| |
Db        129 SQIGFVMNIALNIVTKHVVGRLRPHFLDVCKLANDTCVTGDSHRYITDYTCTGPPELVLE 188

Qy        103 GRKSFPSGHSSFAFAGLAFASFYLAGKLHCFTPQGRGKSWRFCAFLSPLLFAAV---IAL 159
               |||| ||||: :     ::: |:  :|       |   : |    :|  |   :   |:
Db        189 ARKSFYSGHSAVSLYCATWSALYIQARL------GPVLNNRIVVPISQTLMFMIGLGISF 242

Qy        160 SRTCDYKHHWQDLL 173
              ||  | |||| |:|
Db        243 SRITDNKHHWSDVL 256


RESULT 3
YG1P_YEAST
ID   YG1P_YEAST     STANDARD;      PRT;   239 AA.
AC   P53223;
DT   01-OCT-1996 (Rel. 34, Created)
DT   01-OCT-1996 (Rel. 34, Last sequence update)
DT   01-NOV-1997 (Rel. 35, Last annotation update)
DE   Hypothetical 27.6 kDa protein in RPL26B-ACB1 intergenic region.
GN   YGR036C.
OS   Saccharomyces cerevisiae (Baker's yeast).
OC   Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC   Saccharomycetales; Saccharomycetaceae; Saccharomyces.
OX   NCBI_TaxID=4932;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=S288c;
RX   MEDLINE=97435481; PubMed=9290212;
RA   Rieger M., Brueckner M., Schaefer M., Mueller-Auer S.;
RT   "Sequence analysis of 203 kilobases from Saccharomyces cerevisiae
RT   chromosome VII.";
RL   Yeast 13:1077-1090(1997).
CC   -!- SUBCELLULAR LOCATION: Integral membrane protein (Potential).
CC   ----------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ----------------------------------------------------------------------
DR   EMBL; Z72821; CAA97024.1; -.
DR   PIR; S64327; S64327.
DR   SGD; S0003268; CAX4.
DR   GO; GO:0030176; C:endoplasmic reticulum membrane, intrinsic p. . .; IDA.
DR   GO; GO:0016462; F:pyrophosphatase activity; IDA.
DR   GO; GO:0008610; P:lipid biosynthesis; IGI.
DR   GO; GO:0006487; P:N-linked glycosylation; IMP.
DR   InterPro; IPR000326; PA_PTPase.
DR   Pfam; PF01569; PAP2; 1.
DR   SMART; SM00014; acidPPc; 1.
KW   Hypothetical protein; Transmembrane.
FT   TRANSMEM     35     55       POTENTIAL.
FT   TRANSMEM    132    152       POTENTIAL.
FT   TRANSMEM    165    185       POTENTIAL.
SQ   SEQUENCE   239 AA;  27649 MW;  C8EC49CBFC08F8F8 CRC64;

  Query Match             12.8%;  Score 124;  DB 1;  Length 239;
  Best Local Similarity   27.8%;  Pred. No. 8.1e-06;
  Matches   42;  Conservative   21;  Mismatches   50;  Indels   38;  Gaps    8;

Qy         13 YFPTKPMFVIAFLSPLSLIFLAKFLKKADTRDSRQACLAASLALALNGVFTNTIKLIVGR 72
              ||   |: |:||       :|: |:    ||:  :||: |     :| :| | || |: :
Db         36 YFSLMPILVLAF-------YLSWFI---ITRE-LEACIVA-FGQLMNEIFNNVIKNIIKQ 83

Qy         73 PRPDFFYRCFPDGLAHSDLMCTGDKDVVNEGRKSFPSGHSSFAFAGLAFASFYLAGKLHC 132
              ||| |   |                 | :  |    || || |    :|   | : |::
Db         84 PRPVSFGASF------------QNDTIRSG-YGMPSAHSQF----MGFCFTYNSLKIY- 124

Qy        133 FTPQGRGKSWRFCAFLSPLLFAAVIALSRTC 163
                 ||:   ||   :|:  :||   |
Db        125 -------TSWKNLNFLEKCIFSGALALLSFC 148


RESULT 4
Y374_METJA
ID   Y374_METJA     STANDARD;      PRT;   330 AA.
AC   Q57819;
DT   01-NOV-1997 (Rel. 35, Created)
DT   01-NOV-1997 (Rel. 35, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   Hypothetical protein MJ0374.
GN   MJ0374.
OS   Methanococcus jannaschii.
OC   Archaea; Euryarchaeota; Methanococci; Methanococcales;
OC   Methanocaldococcaceae; Methanocaldococcus.
OX   NCBI_TaxID=2190;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=JAL-1 / DSM 2661 / ATCC 43067;
RX   MEDLINE=96337999; PubMed=8688087;
RA   Bult C.J., White O., Olsen G.J., Zhou L., Fleischmann R.D.,
RA   Sutton G.G., Blake J.A., FitzGerald L.M., Clayton R.A., Gocayne J.D.,
RA   Kerlavage A.R., Dougherty B.A., Tomb J.-F., Adams M.D., Reich C.I.,
RA   Overbeek R., Kirkness E.F., Weinstock K.G., Merrick J.M., Glodek A.,
RA   Scott J.L., Geoghagen N.S.M., Weidman J.F., Fuhrmann J.L., Nguyen D.,
RA   Utterback T.R., Kelley J.M., Peterson J.D., Sadow P.W., Hanna M.C.,
RA   Cotton M.D., Roberts K.M., Hurst M.A., Kaine B.P., Borodovsky M.,
RA   Klenk H.-P., Fraser C.M., Smith H.O., Woese C.R., Venter J.C.;
RT   "Complete genome sequence of the methanogenic archaeon, Methanococcus
RT   jannaschii.";
RL   Science 273:1058-1073(1996).
CC   -!- SIMILARITY: TO M.JANNASCHII MJ0201 AND TO H.INFLUENZAE AND
CC       H.SOMNUS HI0703.
CC   ----------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ----------------------------------------------------------------------
DR   EMBL; U67490; AAB98363.1; -.
DR   PIR; F64346; F64346.
DR   TIGR; MJ0374; -.
DR   InterPro; IPR000252; DedA.
DR   InterPro; IPR000326; PA_PTPase.
DR   Pfam; PF00597; DedA; 1.
DR   Pfam; PF01569; PAP2; 1.
DR   SMART; SM00014; acidPPc; 1.
KW   Hypothetical protein; Complete proteome.
SQ   SEQUENCE   330 AA;  37018 MW;  BA84B0170694C097 CRC64;
```

```
  Query Match             9.3%;  Score 90.5;  DB 1;  Length 330;
  Best Local Similarity   28.0%;  Pred. No. 0.042;
  Matches   47;  Conservative   21;  Mismatches   51;  Indels   49;  Gaps    9;

Qy          7 PYVEAEYFPTKPMFVIAFLSPLSLIFLAKFLKKADTRDSRQACLAASLALALNGVFTNTI 66
              |  :    :|     :||  | :  |   ||  |           | :| ||    |  ::
Db        187 PIISKTAYP-----LIAITSLIIFIKNRKFGMK----------LIFALFLAFMIAF--SL 229

Qy         67 KLIVGRPRPDFFYRCFPDGLAHSDLMCTGDKDVVNEGRK-SFPSGHSSFAFAGLAFASFY 125
              | :|  |||          | :  |:|       ||| : ||||||:: ||  || :  :
Db        230 KYLVNEPRPYLV-------LDNVHLLC-------NEGNEPSFPSGHTTLAFT-LATSLLF 274

Qy        126 LAGKLHCFTPQGRGKSWRFCAFLSPLLFAAVIALSRTCDYKHHWQDLL 173
               : ||                |||   :| ::| ||     |:  |:|
Db        275 YSKKLGIL-------------FLS---WAIIVAYSRVYVGVHYPLDVL 306


RESULT 5
YF69_AQUAE
ID   YF69_AQUAE     STANDARD;      PRT;   378 AA.
AC   O67513;
DT   16-OCT-2001 (Rel. 40, Created)
DT   16-OCT-2001 (Rel. 40, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   Hypothetical protein AQ_1569.
GN   AQ_1569.
OS   Aquifex aeolicus.
OC   Bacteria; Aquificae; Aquificales; Aquificaceae; Aquifex.
OX   NCBI_TaxID=63363;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=VF5;
RX   MEDLINE=98196666; PubMed=9537320;
RA   Deckert G., Warren P.V., Gaasterland T., Young W.G., Lenox A.L.,
RA   Graham D.E., Overbeek R., Snead M.A., Keller M., Aujay M., Huber R.,
RA   Feldman R.A., Short J.M., Olson G.J., Swanson R.V.;
RT   "The complete genome of the hyperthermophilic bacterium Aquifex
RT   aeolicus.";
RL   Nature 392:353-358(1998).
CC   -!- SUBCELLULAR LOCATION: Integral membrane protein (Potential).
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; AE000746; AAC07485.1; -.
DR   PIR; G70435; G70435.
KW   Hypothetical protein; Transmembrane; Complete proteome.
FT   TRANSMEM     12     34       POTENTIAL.
FT   TRANSMEM     44     66       POTENTIAL.
FT   TRANSMEM     92    112       POTENTIAL.
FT   TRANSMEM    132    154       POTENTIAL.
FT   TRANSMEM    161    180       POTENTIAL.
FT   TRANSMEM    200    222       POTENTIAL.
FT   TRANSMEM    235    257       POTENTIAL.
FT   TRANSMEM    267    285       POTENTIAL.
FT   TRANSMEM    290    312       POTENTIAL.
FT   TRANSMEM    327    349       POTENTIAL.
FT   TRANSMEM    356    373       POTENTIAL.
SQ   SEQUENCE   378 AA;  42391 MW;  C6C65473872BDCCB CRC64;

  Query Match             8.0%;  Score 77.5;  DB 1;  Length 378;
  Best Local Similarity   30.1%;  Pred. No. 1.2;
  Matches   41;  Conservative   18;  Mismatches   56;  Indels   21;  Gaps    7;

Qy          9 VEAEYFP---TKPMFVIAF-----LSPLSLIFLAKFL-KKADTRDSRQACLAASLALALN 59
              | : :||    || : |  :     :| |:   || || | ||     |:     :|   |
Db         29 VVSTFFPLFITKHIDVKIYSLSYGISLLASFALAIFLGKLADENALRKKFF--TLFSLLT 86


Qy         60 GVFTNTIKLIVGRPRPDFFYRCFPDGL----AHSDLMCTGDKDVVNEGRKSFPSGHS-SF 114
              :| ::| |: | |      |      |    :|   |   :  ::|   | | ||   ||
Db         87 SLFLSSIALLYGTP-----YLALLSFLLMLISHQQAMVFYNSLLLNFENKGFASGLGVSF 141

Qy        115 AFAGLAFASFYLAGKL 130
               : | | |  :|| ||
Db        142 GYVGSAIALIFLADKL 157


RESULT 6
P_HUMAN
ID   P_HUMAN        STANDARD;      PRT;   838 AA.
AC   Q04671; Q15212;
DT   01-JUN-1994 (Rel. 29, Created)
DT   01-JUN-1994 (Rel. 29, Last sequence update)
DT   15-SEP-2003 (Rel. 42, Last annotation update)
DE   P protein (Melanocyte-specific transporter protein).
GN   OCA2 OR P.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=93133287; PubMed=8421497;
RA   Rinchik E.M., Bultman S.J., Horsthemke B., Lee S.-T., Strunk K.M.,
RA   Spritz R.A., Avidano K.A., Jong M.T.C., Nicholls R.D.;
RT   "A gene for the mouse pink-eyed dilution locus and for human type II
RT   oculocutaneous albinism.";
RL   Nature 361:72-76(1993).
RN   [2]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=95324928; PubMed=7601462;
RA   Lee S.-T., Nicholls R.D., Jong M.T.C., Fukai K., Spritz R.A.;
RT   "Organization and sequence of the human P gene and identification of
RT   a new family of transport proteins.";
RL   Genomics 26:354-363(1995).
RN   [3]
RP   REVIEW ON OCA-II VARIANTS.
RX   MEDLINE=99140254; PubMed=10094567;
RA   Oetting W.S., King R.A.;
RT   "Molecular basis of albinism: mutations and polymorphisms of
RT   pigmentation genes associated with albinism.";
RL   Hum. Mutat. 13:99-115(1999).
RN   [4]
RP   VARIANTS OCA-II.
RX   MEDLINE=95179125; PubMed=7874125;
RA   Lee S.-T., Nicholls R.D., Schnur R.E., Guida L.C., Lu-Kuo J.,
RA   Spinner N.B., Zackai E.H., Spritz R.A.;
RT   "Diverse mutations of the P gene among African-Americans with type II
RT   (tyrosinase-positive) oculocutaneous albinism (OCA2).";
RL   Hum. Mol. Genet. 3:2047-2051(1994).
RN   [5]
RP   VARIANTS OCA-II.
RX   MEDLINE=95282778; PubMed=7762554;
RA   Spritz R.A., Fukai K., Holmes S.A., Luande J.;
RT   "Frequent intragenic deletion of the P gene in Tanzanian patients
RT   with type II oculocutaneous albinism (OCA2).";
RL   Am. J. Hum. Genet. 56:1320-1323(1995).
RN   [6]
RP   VARIANTS OCA-II.
RX   MEDLINE=97403942; PubMed=9259203;
RA   Spritz R.A., Lee S.-T., Fukai K., Brondum-Nielsen K., Chitayat D.,
RA   Lipson M.H., Musarella M.A., Rosenmann A., Weleber R.G.;
RT   "Novel mutations of the P gene in type II oculocutaneous albinism
RT   (OCA2).";
RL   Hum. Mutat. 10:175-177(1997).
RN   [7]
RP   VARIANTS OCA-II ARG-86; PHE-112; VAL-368; ILE-592; PRO-724 AND
RP   VAL-787, AND VARIANT ALA-257.
RA   Oetting W.S., Gardner J.M., Fryer J.P., Ching A., Durham-Pierre D.,
RA   King R.A., Brilliant M.H.;
RT   "Mutations of the human P gene associated with type II oculocutaneous
```

```
RT   albinism (OCA2).";
RL   Hum. Mutat. 12:434-434(1998).
CC   -!- FUNCTION: MAY BE INVOLVED IN THE TRANSPORT OF TYROSINE.
CC   -!- SUBCELLULAR LOCATION: Integral membrane protein.
CC   -!- DISEASE: DEFECTS IN OCA2 ARE THE CAUSE OF TYROSINASE-POSITIVE
CC       OCULOCUTANEOUS ALBINISM (OCA-II OR OCA2). OCA2 IS AN AUTOSOMAL
CC       RECESSIVE DISORDER OF PIGMENTATION IN THE SKIN, HAIR, AND EYES.
CC       THE PHENOTYPE OF PATIENTS WITH OCA2 IS TYPICALLY SOMEWHAT LESS
CC       SEVERE THAN IN THOSE WITH TYROSINASE-DEFICIENT OCA1. THERE ARE
CC       SEVERAL FORMS OF OCA2, FROM TYPICAL OCA TO RELATIVELY MILD
CC       "AUTOSOMAL RECESSIVE OCULAR ALBINISM" (AROA). OCA2 IS THE MOST
CC       PREVALENT TYPE OF ALBINISM THROUGHOUT THE WORLD.
CC   -!- SIMILARITY: BELONGS TO THE SLC13A FAMILY OF TRANSPORTERS.
CC       P SUBFAMILY.
CC   -!- DATABASE: NAME=Albinism database (ADB);
CC       NOTE=OCA-II mutations;
CC       WWW="http://www.cbc.umn.edu/tad/".
CC   -!- DATABASE: NAME=Mutations of the P gene;
CC       NOTE=Retina International's Scientific Newsletter;
CC       WWW="http://www.retina-international.com/sci-news/pgenemut.htm".
CC   ----------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ----------------------------------------------------------------------
DR   EMBL; M99564; AAA36477.1; -.
DR   EMBL; U19176; AAC13784.1; -.
DR   EMBL; U19153; AAC13784.1; JOINED.
DR   EMBL; U19154; AAC13784.1; JOINED.
DR   EMBL; U19155; AAC13784.1; JOINED.
DR   EMBL; U19157; AAC13784.1; JOINED.
DR   EMBL; U19158; AAC13784.1; JOINED.
DR   EMBL; U19159; AAC13784.1; JOINED.
DR   EMBL; U19160; AAC13784.1; JOINED.
DR   EMBL; U19161; AAC13784.1; JOINED.
DR   EMBL; U19162; AAC13784.1; JOINED.
DR   EMBL; U19163; AAC13784.1; JOINED.
DR   EMBL; U19164; AAC13784.1; JOINED.
DR   EMBL; U19165; AAC13784.1; JOINED.
DR   EMBL; U19166; AAC13784.1; JOINED.
DR   EMBL; U19167; AAC13784.1; JOINED.
DR   EMBL; U19168; AAC13784.1; JOINED.
DR   EMBL; U19169; AAC13784.1; JOINED.
DR   EMBL; U19171; AAC13784.1; JOINED.
DR   EMBL; U19172; AAC13784.1; JOINED.
DR   EMBL; U19173; AAC13784.1; JOINED.
DR   EMBL; U19174; AAC13784.1; JOINED.
DR   EMBL; U19175; AAC13784.1; JOINED.
DR   PIR; S28911; S28911.
DR   Genew; HGNC:8101; OCA2.
DR   MIM; 203200; -.
DR   MIM; 203310; -.
DR   GO; GO:0005737; C:cytoplasm; TAS.
DR   GO; GO:0016021; C:integral to membrane; TAS.
DR   GO; GO:0005215; F:transporter activity; TAS.
DR   GO; GO:0005302; F:tyrosine transporter activity; TAS.
DR   GO; GO:0006726; P:eye pigment biosynthesis; TAS.
DR   InterPro; IPR000802; Ars_pump.
DR   TIGRFAMs; TIGR00935; 2a45; 1.
KW   Transport; Transmembrane; Albinism; Disease mutation; Polymorphism.
FT   DOMAIN        1    179       CYTOPLASMIC (POTENTIAL).
FT   TRANSMEM    180    197       POTENTIAL.
FT   DOMAIN      198    330       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM    331    347       POTENTIAL.
FT   DOMAIN      348    353       CYTOPLASMIC (POTENTIAL).
FT   TRANSMEM    354    370       POTENTIAL.
FT   DOMAIN      371    384       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM    385    401       POTENTIAL.
FT   DOMAIN      402    423       CYTOPLASMIC (POTENTIAL).
FT   TRANSMEM    424    440       POTENTIAL.
FT   DOMAIN      441    513       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM    514    530       POTENTIAL.
FT   DOMAIN      531    620       CYTOPLASMIC (POTENTIAL).
FT   TRANSMEM    621    637       POTENTIAL.
FT   DOMAIN      638    647       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM    648    664       POTENTIAL.
FT   DOMAIN      665    679       CYTOPLASMIC (POTENTIAL).
FT   TRANSMEM    680    696       POTENTIAL.
FT   DOMAIN      697    720       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM    721    737       POTENTIAL.
FT   DOMAIN      738    760       CYTOPLASMIC (POTENTIAL).
FT   TRANSMEM    761    777       POTENTIAL.
FT   DOMAIN      778    817       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM    818    834       POTENTIAL.
FT   DOMAIN      835    838       CYTOPLASMIC (POTENTIAL).
FT   CARBOHYD    214    214       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    218    218       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    273    273       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    442    442       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    781    781       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   VARIANT      27     27       G -> R (IN OCA-II).
FT                                /FTId=VAR_006117.
FT   VARIANT      86     86       S -> R (IN OCA-II).
FT                                /FTId=VAR_006118.
FT   VARIANT     112    112       C -> F (IN OCA-II).
FT                                /FTId=VAR_006119.
FT   VARIANT     206    211       MISSING (IN OCA-II; SEVERE).
FT                                /FTId=VAR_006120.
FT   VARIANT     257    257       D -> A.
FT                                /FTId=VAR_006121.
FT   VARIANT     273    274       NW -> KV (IN OCA-II).
FT                                /FTId=VAR_006122.
FT   VARIANT     305    305       R -> W (IN OCA-II).
FT                                /FTId=VAR_006123.
FT   VARIANT     368    368       A -> V (IN OCA-II).
FT                                /FTId=VAR_006124.
FT   VARIANT     385    385       F -> I (IN OCA-II; SEVERE).
FT                                /FTId=VAR_006125.
FT   VARIANT     395    395       M -> L (IN OCA-II; SEVERE).
FT                                /FTId=VAR_006126.
FT   VARIANT     404    404       T -> M (IN OCA-II).
FT                                /FTId=VAR_006127.
FT   VARIANT     419    419       R -> Q.
FT                                /FTId=VAR_006128.
FT   VARIANT     419    419       R -> W (IN OCA-II).
FT                                /FTId=VAR_006129.
FT   VARIANT     425    425       MISSING (IN OCA-II; MILD).
FT                                /FTId=VAR_006130.
FT   VARIANT     440    440       L -> F.
FT                                /FTId=VAR_007939.
FT   VARIANT     440    440       L -> H.
FT                                /FTId=VAR_006131.
FT   VARIANT     443    443       V -> I (IN OCA-II).
FT                                /FTId=VAR_006132.
FT   VARIANT     446    446       M -> V (IN MILD OCA-II; AROA FORM).
FT                                /FTId=VAR_006133.
FT   VARIANT     473    473       I -> S (IN OCA-II).
FT                                /FTId=VAR_006134.
FT   VARIANT     481    481       A -> T (IN OCA-II).
FT                                /FTId=VAR_007940.
FT   VARIANT     489    489       N -> D (IN MILD/OCA-II; SEVERE).
FT                                /FTId=VAR_006135.
FT   VARIANT     549    549       H -> Q (IN OCA-II).
FT                                /FTId=VAR_006136.

  Query Match             7.9%;  Score 76.5;  DB 1;  Length 838;
  Best Local Similarity   25.2%;  Pred. No. 3.7;
  Matches   36;  Conservative   17;  Mismatches   57;  Indels   33;  Gaps    5;

Qy         43 RDSRQACLAASLALALNGVFTNTIKLIVGRPRPDFFYRCFPDGLAHSDLMCTGDKDVVNE 102
              || |:   | :: |    | :   :|: |: |        | |   :|
Db          6 RDGRRYPGAPAVELLQTSVPSGLAELVAGKRR-------LPRGAGGAD------------ 46
```

```
Qy        103 GRKSFP---SGHSSFAFAGLAFASFYLAGKLHCFTPQGRGKSWRFCAFL--SPLLFAAVI 157
                  | |   :| ||:| ||  ||||   |: |   ||      :   |   :|||  ::
Db         47 PSHSCPRGAAGQSSWAPAGQEFASFLTKGRSHSSLPQMSSSRSKDSCFTENTPLLRNSLQ 106

Qy        158 ALSRTCDYKHH---------WQD 171
                |   :|          |:|
Db        107 EKGSRCIPVYHPEFITAEESWED 129


RESULT 7
CRNA_EMENI
ID   CRNA_EMENI     STANDARD;      PRT;   507 AA.
AC   P22152;
DT   01-AUG-1991 (Rel. 19, Created)
DT   01-FEB-1996 (Rel. 33, Last sequence update)
DT   01-OCT-1996 (Rel. 34, Last annotation update)
DE   Nitrate transporter (Nitrate permease).
GN   CRNA.
OS   Emericella nidulans (Aspergillus nidulans).
OC   Eukaryota; Fungi; Ascomycota; Pezizomycotina; Eurotiomycetes;
OC   Eurotiales; Trichocomaceae; Emericella.
OX   NCBI_TaxID=162425;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=91095428; PubMed=1986367;
RA   Unkles S.E., Hawker K.L., Grieve C., Campbell E.I., Montague P.,
RA   Kinghorn J.R.;
RT   "crnA encodes a nitrate transporter in Aspergillus nidulans.";
RL   Proc. Natl. Acad. Sci. U.S.A. 88:204-208(1991).
RN   [2]
RP   ERRATUM.
RA   Unkles S.E., Hawker K.L., Grieve C., Campbell E.I., Montague P.,
RA   Kinghorn J.R.;
RL   Proc. Natl. Acad. Sci. U.S.A. 88:4564-4564(1991).
CC   -!- FUNCTION: PERMEASE FOR NITRATE UPTAKE.
CC   -!- SUBCELLULAR LOCATION: Integral membrane protein.
CC   -!- INDUCTION: SUBJECT TO NITRATE AND NITRITE INDUCTION, AND NITROGEN
CC       METABOLITE REPRESSION. CRNA EXPRESSION IS MEDIATED BY THE PRODUCTS
CC       OF NIRA, AREA, AND NIAD.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; M61125; AAA62125.1; -.
DR   EMBL; U34382; AAA76713.1; -.
DR   InterPro; IPR007114; MFS.
DR   InterPro; IPR004737; NO3_transporter.
DR   InterPro; IPR005828; Sub_transporter.
DR   Pfam; PF00083; sugar_tr; 1.
DR   TIGRFAMs; TIGR00886; 2A0108; 1.
KW   Nitrate assimilation; Transport; Transmembrane.
FT   DOMAIN        1     34       CYTOPLASMIC (POTENTIAL).
FT   TRANSMEM     35     55       POTENTIAL.
FT   DOMAIN       56     71       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM     72     92       POTENTIAL.
FT   DOMAIN       93    100       CYTOPLASMIC (POTENTIAL).
FT   TRANSMEM    101    121       POTENTIAL.
FT   DOMAIN      122    130       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM    131    151       POTENTIAL.
FT   DOMAIN      152    161       CYTOPLASMIC (POTENTIAL).
FT   TRANSMEM    162    182       POTENTIAL.
FT   DOMAIN      183    198       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM    199    219       POTENTIAL.
FT   DOMAIN      220    306       CYTOPLASMIC (HYDROPHILIC) (POTENTIAL).
FT   TRANSMEM    307    327       POTENTIAL.
FT   DOMAIN      328    357       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM    358    378       POTENTIAL.
FT   DOMAIN      379    389       CYTOPLASMIC (POTENTIAL).
FT   TRANSMEM    390    410       POTENTIAL.
FT   DOMAIN      411    417       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM    418    438       POTENTIAL.
FT   DOMAIN      439    507       CYTOPLASMIC (POTENTIAL).
SQ   SEQUENCE   507 AA;  54925 MW;  4A3D3FA643F16952 CRC64;

  Query Match             7.7%;  Score 74.5;  DB 1;  Length 507;
  Best Local Similarity   22.8%;  Pred. No. 3.4;
  Matches   49;  Conservative   31;  Mismatches   64;  Indels   71;  Gaps   12;

Qy         20 FVIAFLS-----PLSLIFLAKFLKKADTRDS------------RQAC------------ 49
              |::||||     ||  : :   |  : |: :                |  |
Db         43 FMLAFLSWYAFPPLLTVTIRDDLDMSQTQIANSNIIALLATLLVRLICGPLCDRFGPRLV 102

Qy         50 -----LAASLALALNGVFTNTIKLIVGRPRPDFFYRCFPDGLAHSDLMCTG--DKDVVNE 102
                   |  |:  |: |:  |:   ||  |    ||             : |||  || :|
Db        103 FIGLLLVGSIPTAMAGLVTSPQGLIALR----FFIGILGGTFVPCQVWCTGFFDKSIV-- 156

Qy        103 GRKSFPSGHSSFAFAGLAFA----SFYLAGKLHCFTPQGRG----KSWRFCAFLSP--LL 152
                  | ::   ||| |    ::::   :    : :|    |:||  |:: |  |:
Db        157 -------GTANSLAAGLGNAGGGITYFVMPAIFDSLIRDQGLPAHKAWR-VAYIVPFILI 208

Qy        153 FAAVIALSRTCD--------YKHHWQDLLKCTNTA 179
                || : :  |||         :| |  : : | ||
Db        209 VAAALGMLFTCDDTPTGKWSERHIW--MKEDTQTA 241


RESULT 8
ITH4_PIG
ID   ITH4_PIG       STANDARD;      PRT;   921 AA.
AC   P79263;
DT   15-JUL-1998 (Rel. 36, Created)
DT   15-JUL-1998 (Rel. 36, Last sequence update)
DT   28-FEB-2003 (Rel. 41, Last annotation update)
DE   Inter-alpha-trypsin inhibitor heavy chain H4 precursor (ITI heavy
DE   chain H4) (Inter-alpha-inhibitor heavy chain 4) (Inter-alpha-trypsin
DE   inhibitor family heavy chain-related protein) (IHRP) (Major acute
DE   phase protein) (MAP).
GN   ITIH4 OR IHRP.
OS   Sus scrofa (Pig).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.
OX   NCBI_TaxID=9823;
RN   [1]
RP   SEQUENCE FROM N.A., AND SEQUENCE OF 28-36; 695-703 AND 710-722.
RC   TISSUE=Liver;
RX   MEDLINE=96271024; PubMed=8830057;
RA   Hashimoto K., Tobe T., Sumiya J.-I., Sano Y., Choi-Miura N.-H.,
RA   Ozawa A., Yasue H., Tomita M.;
RT   "Primary structure of the pig homologue of human IHRP: inter-alpha-
RT   trypsin inhibitor family heavy chain-related protein.";
RL   J. Biochem. 119:577-584(1996).
RN   [2]
RP   PRELIMINARY SEQUENCE OF 267-556 FROM N.A.
RC   TISSUE=Liver;
RX   MEDLINE=90371455; PubMed=1697703;
RA   Buchman T.G., Cabin D.E., Vickers S., Deutschman C.S., Delgado E.,
RA   Sussman M.M., Bulkley G.B.;
RT   "Molecular biology of circulatory shock. Part II. Expression of four
RT   groups of hepatic genes is enhanced after resuscitation from
RT   cardiogenic shock.";
RL   Surgery 108:559-566(1990).
RN   [3]
RP   SEQUENCE OF 28-54 AND 223-240.
RC   TISSUE=Serum;
RX   MEDLINE=96013138; PubMed=7556597;
RA   Gonzalez-Ramon N., Alava M.A., Sarsa J.A., Pineiro M., Escartin A.,
RA   Garcia-Gil A., Lampreave F., Pineiro A.;
RT   "The major acute phase serum protein in pigs is homologous to human
RT   plasma kallikrein sensitive PK-120.";
RL   FEBS Lett. 371:227-230(1995).
CC   -!- FUNCTION: MAY BE INVOLVED IN ACUTE PHASE REACTIONS.
```

```
CC   -!- TISSUE SPECIFICITY: Liver-specific.
CC   -!- INDUCTION: LEVELS INCREASE SIGNIFICANTLY AFTER CARDIOGENIC
CC       SHOCK.
CC   -!- PTM: APPEARS TO BE BOTH N- AND O-GLYCOSYLATED (BY SIMILARITY).
CC   -!- PTM: CLEAVED BY PLASMA KALLIKREIN TO YIELD 55- AND 25-KDA
CC       FRAGMENTS.
CC   -!- SIMILARITY: BELONGS TO THE ITIH FAMILY.
CC   -!- SIMILARITY: Contains 1 VWFA domain.
CC   -!- CAUTION: REF.2 SEQUENCE IS INCORRECT DUE TO FRAMESHIFTS AND OTHER
CC       SEQUENCING ERRORS.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; U43164; AAD00024.1; -.
DR   EMBL; S82800; AAB46821.1; -.
DR   EMBL; M29507; -; NOT_ANNOTATED_CDS.
DR   PIR; JC4625; JC4625.
DR   InterPro; IPR006587; VIT.
DR   InterPro; IPR002035; VWF_A.
DR   Pfam; PF00092; vwa; 1.
DR   SMART; SM00609; VIT; 1.
DR   SMART; SM00327; VWA; 1.
DR   PROSITE; PS50234; VWFA; 1.
KW   Serine protease inhibitor; Repeat; Signal; Multigene family;
KW   Glycoprotein.
FT   SIGNAL       1     27
FT   CHAIN       28    921       INTER-ALPHA-TRYPSIN INHIBITOR HEAVY CHAIN
FT                               H4.
FT   DOMAIN     270    428       VWFA.
FT   CARBOHYD    80     80       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   205    205       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   242    242       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   513    513       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   577    577       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CONFLICT    49     50       HT -> SK (IN REF. 3).
FT   CONFLICT   703    703       D -> H (IN REF. 1; AA SEQUENCE).
SQ   SEQUENCE   921 AA;  102146 MW;  E2BF95925DE8D07C CRC64;

  Query Match             7.5%;  Score 73;  DB 1;  Length 921;
  Best Local Similarity   25.3%;  Pred. No. 9.7;
  Matches   40;  Conservative   30;  Mismatches   56;  Indels   32;  Gaps    8;

Qy         10 EAEYFPTKPMF------VIAFLSPLSLIFLAKFLKKADTRDSRQACLAASLALALNGVF- 62
              |||:   | :|      : |:|:   |  ||: :  :|    ::|  | :|:|:||  |
Db        526 EAEFLSPKYIFHSFMERLWAYLTIQQL--LAQTVSASDA--EKKALEARALSLSLNYSFV 581

Qy         63 TNTIKLIVGRPRPDFFYRCFPDGLAHSDLMCTGDKDVVNEGRK-SFPSGHSSFAFAGLAF 121
                 |    ::: :          |:|   | :    :| | |  |: :  |||||| |  :
Db        582 TPLTSMVITK----------PEGQEQSQV---AEKPVENGNRQGNTHSGHSSFQFHSVGD 628

Qy        122 ASFYLAGK-------LHCFTPQGRGKSWRFCAFLSPLL 152
               :  | |          |    :|: : :   ::| | |
Db        629 RTSRLTGGSSVDPVFSHRRGWKGQAQGFEKMSYLPPRL 666


RESULT 9
ML12_ARATH
ID   ML12_ARATH     STANDARD;      PRT;   576 AA.
AC   O80961; Q94KB3;
DT   15-JUL-1999 (Rel. 38, Created)
DT   28-FEB-2003 (Rel. 41, Last sequence update)
DT   28-FEB-2003 (Rel. 41, Last annotation update)
DE   MLO-like protein 12 (AtMlo12).
GN   MLO12 OR AT2G39200 OR T16B24.16.
OS   Arabidopsis thaliana (Mouse-ear cress).
OC   Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC   Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC   eurosids II; Brassicales; Brassicaceae; Arabidopsis.
OX   NCBI_TaxID=3702;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Devoto A., Hartmann H.A., Piffanelli P., Elliott C., Simmons C.,
RA   Taramino G., Goh C., Cohen F.E., Schulze-Lefert P., Panstruga R.;
RT   "Molecular phylogeny and domain-specific co-evolution of the
RT   plant-specific seven transmembrane MLO family.";
RL   Submitted (APR-2001) to the EMBL/GenBank/DDBJ databases.
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=cv. Columbia;
RX   MEDLINE=20083487; PubMed=10617197;
RA   Lin X., Kaul S., Rounsley S.D., Shea T.P., Benito M.-I., Town C.D.,
RA   Fujii C.Y., Mason T.M., Bowman C.L., Barnstead M.E., Feldblyum T.V.,
RA   Buell C.R., Ketchum K.A., Lee J.J., Ronning C.M., Koo H.L.,
RA   Moffat K.S., Cronin L.A., Shen M., Pai G., Van Aken S., Umayam L.,
RA   Tallon L.J., Gill J.E., Adams M.D., Carrera A.J., Creasy T.H.,
RA   Goodman H.M., Somerville C.R., Copenhaver G.P., Preuss D.,
RA   Nierman W.C., White O., Eisen J.A., Salzberg S.L., Fraser C.M.,
RA   Venter J.C.;
RT   "Sequence and analysis of chromosome 2 of the plant Arabidopsis
RT   thaliana.";
RL   Nature 402:761-768(1999).
CC   -!- SUBCELLULAR LOCATION: Integral membrane protein (Potential).
CC   -!- SIMILARITY: BELONGS TO THE MLO FAMILY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; AF369573; AAK53805.1; -.
DR   EMBL; AC004697; AAC28997.2; -.
DR   InterPro; IPR004326; Mlo.
DR   Pfam; PF03094; Mlo; 1.
KW   Transmembrane; Multigene family.
FT   TRANSMEM     14     34       POTENTIAL.
FT   TRANSMEM     60     80       POTENTIAL.
FT   TRANSMEM    156    176       POTENTIAL.
FT   TRANSMEM    279    299       POTENTIAL.
FT   TRANSMEM    309    329       POTENTIAL.
FT   TRANSMEM    363    383       POTENTIAL.
FT   TRANSMEM    406    426       POTENTIAL.
SQ   SEQUENCE   576 AA;  66548 MW;  43DA9F6AED64D8E6 CRC64;

  Query Match             7.3%;  Score 70.5;  DB 1;  Length 576;
  Best Local Similarity   22.2%;  Pred. No. 10;
  Matches   38;  Conservative   23;  Mismatches   57;  Indels   53;  Gaps    8;

Qy         20 FVIAFLSPL---SLIFLAKFLKKADTRDSRQACLAASLALALNGVFTNTIKLIVGRPRPD 76
              ||: |:| :    | |:  : ||   :   :|       | | | | : : :::  |  :
Db         21 FVLLFISIMIEYFLHFIGHWFKKKHKKALSEALEKVKAELMLLG-FISLLLVVLQTPVSE 79

Qy         77 FFYRCFPDGLA--------HSDLMCTGDKDVVNEGRK--------SFPSGHSSFAFAG-- 118
                 | |  :|     | ::   | || :::|||         | |   : |  |
Db         80 I---CIPRNIAATWHPCSNHQEIAKYG-KDYIDDGRKILEDFDSNDFYSPRRNLATKGYD 135

Qy        119 -------LAFASFYLAGKLHCF-------------------TPQGRGKSW 142
                    :|  | |   :|| |                     |   : |||
Db        136 KCAEKGKVALVSAYGIHQLHIFIFVLAVFHVLYCIITYALGKTKMKKWKSW 186


RESULT 10
E1BS_ADECT
ID   E1BS_ADECT     STANDARD;      PRT;   133 AA.
AC   P14265;
DT   01-JAN-1990 (Rel. 13, Created)
DT   01-JAN-1990 (Rel. 13, Last sequence update)
DT   01-NOV-1997 (Rel. 35, Last annotation update)
```

```
DE   E1B protein, small T-antigen (Early E1B 15 kDa protein).
OS   Canine adenovirus type 2 (strain Toronto A 26-61).
OC   Viruses; dsDNA viruses, no RNA stage; Adenoviridae; Mastadenovirus.
OX   NCBI_TaxID=69152;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=90021176; PubMed=2800332;
RA   Shibata R., Shinagawa M., Iida Y., Tsukiyama T.;
RT   "Nucleotide sequence of E1 region of canine adenovirus type 2.";
RL   Virology 172:460-467(1989).
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; J04368; AAA42471.1; -.
DR   PIR; B34165; WMADN2.
DR   InterPro; IPR002924; AdenoE1B_19kDa.
DR   InterPro; IPR002475; BCL2_family.
DR   Pfam; PF01691; Adeno_E1B_19K; 1.
DR   ProDom; PD004074; AdenoE1B_19kDa; 1.
DR   PROSITE; PS50062; BCL2_FAMILY; 1.
KW   Early protein.
SQ   SEQUENCE   133 AA;  15219 MW;  2C733F250830B504 CRC64;

  Query Match             7.2%;  Score 70;  DB 1;  Length 133;
  Best Local Similarity   35.7%;  Pred. No. 2.2;
  Matches   30;  Conservative    9;  Mismatches   31;  Indels   14;  Gaps    6;

Qy         75 PDFFYR-CFPDGLAHSDLMCTGDKDVVNEGR--KSFPSGHSSFAFAGLAFASFYLAGKLH 131
              | || | |||    | :|::  |:     |||  :  |  |   || ||    | :|
Db         24 PGFFRRWCFP---ALADVV--GNIVEQEEGRFWQILPENH---AFWGLLRRGFTVASFTE 75

Qy        132 CFTP---QGRGKSWRFCAFLSPLL 152
                 |    : ||:   | ||:| ||
Db         76 IITAAQLENRGRQLAFLAFISFLL 99


RESULT 11
UDP2_PIG
ID   UDP2_PIG       STANDARD;      PRT;   507 AA.
AC   P79303;
DT   15-JUL-1999 (Rel. 38, Created)
DT   15-JUL-1999 (Rel. 38, Last sequence update)
DT   15-JUL-1999 (Rel. 38, Last annotation update)
DE   UTP--glucose-1-phosphate uridylyltransferase 2 (EC 2.7.7.9) (UDP-
DE   glucose pyrophosphorylase 2) (UDPGP 2) (UGPase 2).
GN   UGP2.
OS   Sus scrofa (Pig).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.
OX   NCBI_TaxID=9823;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Muscle;
RA   Looft C., Paul S.;
RT   "cDNA sequencing of the porcine UDP glucose pyrophosphorylase
RT   gene.";
RL   Submitted (JUN-1996) to the EMBL/GenBank/DDBJ databases.
CC   -!- FUNCTION: PLAYS A CENTRAL ROLE AS A GLUCOSYL DONOR IN CELLULAR
CC       METABOLIC PATHWAYS.
CC   -!- CATALYTIC ACTIVITY: UTP + alpha-D-glucose 1-phosphate =
CC       diphosphate + UDP-glucose.
CC   -!- SUBUNIT: Homooctamer (By similarity).
CC   -!- SUBCELLULAR LOCATION: Cytoplasmic.
CC   -!- SIMILARITY: BELONGS TO THE EUKARYOTIC UDPGP FAMILY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; X99312; CAA67690.1; -.
DR   InterPro; IPR002618; UDPGP.
DR   Pfam; PF01704; UDPGP; 1.
KW   Transferase; Kinase; Nucleotidyltransferase; Multigene family.
FT   INIT_MET      0      0       BY SIMILARITY.
SQ   SEQUENCE   507 AA;  56821 MW;  327F66D104D53965 CRC64;

  Query Match             7.2%;  Score 70;  DB 1;  Length 507;
  Best Local Similarity   27.6%;  Pred. No. 10;
  Matches   27;  Conservative   14;  Mismatches   27;  Indels   30;  Gaps    6;

Qy         97 KDVVNEGRKS---FPSGHSSFAFAGLAFASFYLAGKLHCFTPQGRGKSWRFCAFLSPLLF 153
              |||   |  :   :| ||       |  :|||| :| |  |    | || : | : :  |
Db        205 KDVSYSGENTEAWYPPGH------GDIYASFYNSGLLDTFI--GEGKEYIFVSNIDNL-- 254

Qy        154 AAVIAL----------SRTCDYKHHWQDLLKCTNTAK 180
              | : |           | |::    ::: || |:
Db        255 GATVDLYILNHLMNPPNGRPCEF------VMEATNKAR 286


RESULT 12
ABC3_HUMAN
ID   ABC3_HUMAN     STANDARD;      PRT;  1704 AA.
AC   Q99758; Q92473;
DT   30-MAY-2000 (Rel. 39, Created)
DT   30-MAY-2000 (Rel. 39, Last sequence update)
DT   28-FEB-2003 (Rel. 41, Last annotation update)
DE   ATP-binding cassette, sub-family A, member 3 (ATP-binding cassette
DE   transporter 3) (ATP-binding cassette 3) (ABC-C transporter).
GN   ABCA3 OR ABC3.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Thyroid carcinoma;
RX   MEDLINE=96326608; PubMed=8706931;
RA   Klugbauer N., Hofmann F.;
RT   "Primary structure of a novel ABC transporter with a chromosomal
RT   localization on the band encoding the multidrug resistance-associated
RT   protein.";
RL   FEBS Lett. 391:61-65(1996).
RN   [2]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=97179225; PubMed=9027511;
RA   Connors T.D., van Raay T.J., Petry L.R., Klinger K.W., Landes G.M.,
RA   Burn T.C.;
RT   "The cloning of a human ABC gene (ABC3) mapping to chromosome
RT   16p13.3.";
RL   Genomics 39:231-234(1997).
CC   -!- FUNCTION: MAY BE A TRANSPORTER, ITS NATURAL SUBSTRATE HAS NOT BEEN
CC       FOUND YET (BY SIMILARITY). MAY ACT AS AN EFFLUX PUMP FOR
CC       CHEMOTHERAPEUTICS DRUGS.
CC   -!- TISSUE SPECIFICITY: HIGHLY EXPRESSED IN LUNG, FOLLOWED BY BRAIN,
CC       PANCREAS, SKELETAL MUSCLE AND HEART. WEAKLY EXPRESSED IN PLACENTA,
CC       KIDNEY AND LIVER. ALSO EXPRESSED IN MEDULLARY THYROID CARCINOMA
CC       CELLS (MTC) AND IN C-CELL CARCINOMA.
CC   -!- DOMAIN: MULTIFUNCTIONAL POLYPEPTIDE WITH TWO HOMOLOGOUS HALVES,
CC       EACH CONTAINING AN HYDROPHOBIC MEMBRANE-ANCHORING DOMAIN AND AN
CC       ATP BINDING CASSETTE (ABC) DOMAIN (BY SIMILARITY).
CC   -!- SIMILARITY: BELONGS TO THE ABC TRANSPORTER FAMILY. ABCA SUBFAMILY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
```

```
CC   modified and this statement is not removed.  Usage  by  and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ----------------------------------------------------------------------
DR   EMBL; U78735; AAC50967.1; -.
DR   EMBL; X97187; CAA65825.1; -.
DR   PIR; A59188; A59188.
DR   PIR; S71363; S71363.
DR   Genew; HGNC:33; ABCA3.
DR   MIM; 601615; -.
DR   GO; GO:0016021; C:integral to membrane; TAS.
DR   GO; GO:0005624; C:membrane fraction; TAS.
DR   GO; GO:0004009; F:ATP-binding cassette (ABC) transporter acti. . .; TAS.
DR   GO; GO:0005215; F:transporter activity; TAS.
DR   GO; GO:0009315; P:drug resistance; TAS.
DR   GO; GO:0006832; P:small molecule transport; TAS.
DR   InterPro; IPR003593; AAA_ATPase.
DR   InterPro; IPR003439; ABC_transporter.
DR   Pfam; PF00005; ABC_tran; 2.
DR   ProDom; PD000006; ABC_transporter; 2.
DR   SMART; SM00382; AAA; 2.
DR   PROSITE; PS00211; ABC_TRANSPORTER_1; 1.
DR   PROSITE; PS50893; ABC_TRANSPORTER_2; 2.
KW   ATP-binding; Transport; Transmembrane.
FT   TRANSMEM     22     42       POTENTIAL.
FT   TRANSMEM    249    269       POTENTIAL.
FT   TRANSMEM    307    327       POTENTIAL.
FT   TRANSMEM    344    364       POTENTIAL.
FT   TRANSMEM    373    393       POTENTIAL.
FT   TRANSMEM    405    425       POTENTIAL.
FT   TRANSMEM    447    467       POTENTIAL.
FT   TRANSMEM    925    945       POTENTIAL.
FT   TRANSMEM   1100   1120       POTENTIAL.
FT   TRANSMEM   1144   1164       POTENTIAL.
FT   TRANSMEM   1183   1203       POTENTIAL.
FT   TRANSMEM   1213   1233       POTENTIAL.
FT   TRANSMEM   1245   1265       POTENTIAL.
FT   TRANSMEM   1306   1326       POTENTIAL.
FT   NP_BIND     566    573       ATP (POTENTIAL).
FT   NP_BIND    1416   1423       ATP (POTENTIAL).
FT   CONFLICT     36     36       P -> S (IN REF. 2).
FT   CONFLICT    196    196       L -> P (IN REF. 2).
SQ   SEQUENCE   1704 AA;  191387 MW;  AF0098DAF7A04F5F CRC64;

  Query Match             7.1%;  Score 69;  DB 1;  Length 1704;
  Best Local Similarity   26.1%;  Pred. No. 52;
  Matches   35;  Conservative   11;  Mismatches   40;  Indels   48;  Gaps    7;

Qy          51 AASLALALNGVFTNTIKLIVGRPRPDFFYRCFPDGLAHSDLMCTGDKDVVNEGRKSF--- 107
               | :||:  | :|    ||: | |         ||     | |     ||  ||||| |
Db        1051 ATALAVVDNLLF----KLLCG-PHASIVVSNFPQ--PRSALQAA--KDQFNEGRKGFDIA 1101

Qy         108 ------------------------PSGHSSFAFAGLAFASFYLAGKLHCFTPQGRGKSWR 143
                                       : |  |  :|:  |||:|:  |           |
Db        1102 LNLLFAMAFLASTFSILAVSERAVQAKHVQFV-SGVHVASFWLSALL----------WD 1149

Qy         144 FCAFLSPLLFAAVI 157
                :|| | |    |:
Db        1150 LISFLIPSLLLLVV 1163


RESULT 13
RRPL_PI3H4
ID   RRPL_PI3H4     STANDARD;      PRT;  2233 AA.
AC   P12577;
DT   01-OCT-1989 (Rel. 12, Created)
DT   01-OCT-1989 (Rel. 12, Last sequence update)
DT   28-FEB-2003 (Rel. 41, Last annotation update)
DE   RNA polymerase beta subunit (EC 2.7.7.48) (Large structural protein)
DE   (L protein).
GN   L.
OS   Human parainfluenza 3 virus (strain NIH 47885).
OC   Viruses; ssRNA negative-strand viruses; Mononegavirales;
OC   Paramyxoviridae; Paramyxovirinae; Respirovirus.
OX   NCBI_TaxID=11217;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=88306242; PubMed=2841798;
RA   Galinski M.S., Mink M.A., Pons M.W.;
RT   "Molecular cloning and sequence analysis of the human parainfluenza 3
RT   virus gene encoding the L protein.";
RL   Virology 165:499-510(1988).
RN   [2]
RP   SEQUENCE OF 1-26 FROM N.A.
RX   MEDLINE=88032139; PubMed=2822598;
RA   Storey D.G., Cote M.-J., Dimock K., Kang C.Y.;
RT   "Nucleotide sequence of the coding and flanking regions of the human
RT   parainfluenza virus 3 hemagglutinin-neuraminidase gene: comparison
RT   with other paramyxoviruses.";
RL   Intervirology 27:69-80(1987).
CC   -!- FUNCTION: PROBABLE COMPONENT OF THE ACTIVE POLYMERASE. IT MAY
CC       FUNCTION IN MRNA SYNTHESIS, CAPPING, METHYLATION AND POLY(A)
CC       SYNTHESIS OF NEWLY SYNTHESIZED VIRAL MRNAS, RNA EDITING OF THE P
CC       GENE TRANSCRIPT, AND PROTEIN KINASE ACTIVITY.
CC   -!- CATALYTIC ACTIVITY: N nucleoside triphosphate = N diphosphate +
CC       {RNA}(N).
CC   -!- SIMILARITY: WITH L PROTEIN OF OTHER PARAMYXOVIRUSES.
CC   ----------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between  the Swiss Institute of Bioinformatics  and the  EMBL outstation -
CC   the European Bioinformatics Institute.  There are no  restrictions on  its
CC   use  by  non-profit  institutions as long  as its content  is  in  no  way
CC   modified and this statement is not removed.  Usage  by  and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ----------------------------------------------------------------------
DR   EMBL; M21649; AAA46854.1; -.
DR   EMBL; M20402; AAA46857.1; -.
DR   PIR; B46451; B46451.
DR   InterPro; IPR007098; RNA_pol_monon.
DR   InterPro; IPR001016; Viral_RNA_pol_L.
DR   Pfam; PF00946; Paramyx_RNA_pol; 1.
KW   Transferase; RNA-directed RNA polymerase.
FT   CONFLICT     26     26       K -> R (IN REF. 2).
SQ   SEQUENCE   2233 AA;  255802 MW;  DBBB8B5DE74B4638 CRC64;

  Query Match             7.1%;  Score 69;  DB 1;  Length 2233;
  Best Local Similarity   32.7%;  Pred. No. 71;
  Matches   16;  Conservative   10;  Mismatches   23;  Indels    0;  Gaps    0;

Qy          76 DFFYRCFPDGLAHSDLMCTGDKDVVNEGRKSFPSGHSSFAFAGLAFASF 124
               | | | : :|::    :|  | :| |: :::| | | ||       |||
Db        1576 DLFMREWLNGVSLEIYICDSDMEVANDRKQAFISRHLSFVCCLAEIASF 1624


RESULT 14
ISPZ_BRAJA
ID   ISPZ_BRAJA     STANDARD;      PRT;   200 AA.
AC   P30961;
DT   01-JUL-1993 (Rel. 26, Created)
DT   28-FEB-2003 (Rel. 41, Last sequence update)
DT   15-SEP-2003 (Rel. 42, Last annotation update)
DE   Probable intracellular septation protein.
GN   ISPZ OR BLL0472.
OS   Bradyrhizobium japonicum.
OC   Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;
OC   Bradyrhizobiaceae; Bradyrhizobium.
OX   NCBI_TaxID=375;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=USDA 110;
RX   MEDLINE=22484998; PubMed=12597275;
RA   Kaneko T., Nakamura Y., Sato S., Minamisawa K., Uchiumi T.,
RA   Sasamoto S., Watanabe A., Idesawa K., Iriguchi M., Kawashima K.,
RA   Kohara M., Matsumoto M., Shimpo S., Tsuruoka H., Wada T., Yamada M.,
RA   Tabata S.;
```

```
RT   "Complete genomic sequence of nitrogen-fixing symbiotic bacterium
RT   Bradyrhizobium japonicum USDA110.";
RL   DNA Res. 9:189-197(2002).
RN   [2]
RP   SEQUENCE OF 11-200 FROM N.A.
RC   STRAIN=110RIF15;
RX   MEDLINE=91210304; PubMed=1850420;
RA   Ramseier T.M., Winteler H.V., Hennecke H.;
RT   "Discovery and sequence analysis of bacterial genes involved in the
RT   biogenesis of c-type cytochromes.";
RL   J. Biol. Chem. 266:7793-7803(1991).
CC   -!- FUNCTION: Involved in cell division; probably involved in
CC       intracellular septation (By similarity).
CC   -!- SUBCELLULAR LOCATION: Integral membrane protein (By similarity).
CC   -!- SIMILARITY: BELONGS TO THE ISPZ FAMILY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; AP005936; BAC45737.1; ALT_INIT.
DR   EMBL; M60874; AAA26197.1; -.
DR   PIR; F39741; F39741.
DR   HAMAP; MF_00189; -; 1.
DR   InterPro; IPR006008; SeptationA.
DR   Pfam; PF04279; IspA; 1.
DR   ProDom; PD016710; SeptationA; 1.
DR   TIGRFAMs; TIGR00997; ispZ; 1.
KW   Cell division; Septation; Transmembrane; Complete proteome.
FT   TRANSMEM     32     52       POTENTIAL.
FT   TRANSMEM     56     76       POTENTIAL.
FT   TRANSMEM     93    113       POTENTIAL.
FT   TRANSMEM    126    146       POTENTIAL.
FT   TRANSMEM    153    173       POTENTIAL.
SQ   SEQUENCE   200 AA;  22221 MW;  E86E2553DAE47F25 CRC64;

  Query Match             7.0%;  Score 68;  DB 1;  Length 200;
  Best Local Similarity   22.0%;  Pred. No. 5.7;
  Matches   37;  Conservative   19;  Mismatches   54;  Indels   58;  Gaps    7;

Qy         18 PMFVIAF-LSPLSLIFL--AKFLKKADTRDSRQACLAASLA-------LALNGVFTNTIK 67
              |:| :|  | || : |   |||   | |      | :|| :|       : :  : |  |
Db          8 PLFKLATELGPLLVFFFVNAKFNLFAATGAFMVAIVAAMIASYVVTRHIPIMAIVTGVIV 67

Qy         68 LIVG------------RPRPDFFYRCFPDGLAHSDLMCTGDKDVVNEGRKSFPSGHSSFA 115
              |: |             : :|   |  |                ::
Db         68 LVFGTLTLVLHDETFIKVKPTIIYGLF----------------------------AAIL 98

Qy        116 FAGLAFASFYLA---GKLHCFTPQGRGKSWRFCAFLSPLLFAAVIALS 160
                || |   ::|    ::   ||||  ||       | || :  |:
Db         99 GGGLLFGRSFIAVMFDQMFNLTPQG----WRILTLRWALFFAGMAVLN 142


RESULT 15
YGFR_ECOLI
ID   YGFR_ECOLI     STANDARD;      PRT;   276 AA.
AC   Q46818;
DT   01-NOV-1997 (Rel. 35, Created)
DT   01-NOV-1997 (Rel. 35, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   Hypothetical protein ygfR.
GN   YGFR OR B2885.
OS   Escherichia coli.
OC   Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC   Enterobacteriaceae; Escherichia.
OX   NCBI_TaxID=562;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=K12 / MG1655;
RX   MEDLINE=97426617; PubMed=9278503;
RA   Blattner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,
RA   Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,
RA   Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,
RA   Mau B., Shao Y.;
RT   "The complete genome sequence of Escherichia coli K-12.";
RL   Science 277:1453-1474(1997).
CC   -!- SIMILARITY: BELONGS TO THE YICO/YIEG/YJCD FAMILY. CORRESPONDS TO
CC       THE C-TERMINAL PART; THE N-TERMINAL PART IS YGFQ.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; U28375; AAA83066.1; -.
DR   EMBL; AE000372; AAC75923.1; -.
DR   PIR; E65072; E65072.
DR   EcoGene; EG13068; ygfR.
DR   InterPro; IPR006043; Xant/urac/vitC.
DR   Pfam; PF00860; xan_ur_permease; 1.
KW   Hypothetical protein; Complete proteome.
SQ   SEQUENCE   276 AA;  28407 MW;  F8750315CC35BB9D CRC64;

  Query Match             7.0%;  Score 68;  DB 1;  Length 276;
  Best Local Similarity   27.6%;  Pred. No. 8.3;
  Matches   32;  Conservative   13;  Mismatches   41;  Indels   30;  Gaps    5;

Qy         50 LAASLALALNGVF--TNTIKLIVGRPRPDFFYRCFPDGLAHSDLMCTGDKDVVNEGRKSF 107
              | : ||| :  ||  | ||: : |:                    :   |  ::| | |:
Db         85 LPSVLALVMTAVFDATGTIRAVAGQAN----------------LLDKDNQIIN-GGKAL 126

Qy        108 PSGHSSFAFAGL---AFASFYL-------AGKLHCFTPQGRGKSWRFCAFLSPLLF 153
              |   |  |:||   | |: |:       ||     |    |  :    ||||| |
Db        127 TSDSVSSIFSGLVGAAPAAVYIESAAGTAAGGKTGLTATVVGALFLLILFLSPLSF 182


Search completed: August 14, 2003, 13:04:09
Job time : 24 secs
```